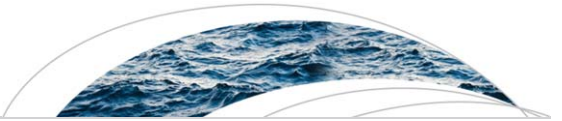
AGU PUBLICATIONS

Water Resources Research



RESEARCH ARTICLE

10.1002/2015WR018247

Global-scale regionalization of hydrologic model parameters

Hylke E. Beck[1], Albert I. J. M. van Dijk[2], Ad de Roo[1], Diego G. Miralles[3,4], Tim R. McVicar[5,6], Jaap Schellekens[7], and L. Adrian Bruijnzeel[8]

[1]European Commission, Joint Research Centre, Ispra, Italy, [2]Fenner School of Environment and Society, Australian National University, Canberra, Australian Capital Territory, Australia, [3]Department of Earth Sciences, Vrije Universiteit Amsterdam, Netherlands, [4]Laboratory of Hydrology and Water Management, Ghent University, Belgium, [5]CSIRO Land and Water, Canberra, Australian Capital Territory, Australia, [6]Australian Research Council Centre of Excellence for Climate System Science, Sydney, Australia, [7]Inland Water Systems Unit, Deltares, Delft, Netherlands, [8]Department of Geography, King's College London, UK

Key Points:
- For the first time, a scheme for regionalization of model parameters at global scale was developed
- The improvement in model performance was about half of that achieved through calibration
- The regionalized model outperformed nine state-of-the-art models including their ensemble mean

Supporting Information:
- Supporting Information S1

Correspondence to:
H. E. Beck,
hylke.beck@jrc.ec.europa.eu

Citation:
Beck, H. E., A. I. J. M. van Dijk, A. de Roo, D. G. Miralles, T. R. McVicar, J. Schellekens, and L. A. Bruijnzeel (2016), Global-scale regionalization of hydrologic model parameters, *Water Resour. Res.*, *52*, doi:10.1002/2015WR018247.

Received 16 OCT 2015
Accepted 14 APR 2016
Accepted article online 20 APR 2016

Abstract Current state-of-the-art models typically applied at continental to global scales (hereafter called macroscale) tend to use a priori parameters, resulting in suboptimal streamflow (*Q*) simulation. For the first time, a scheme for regionalization of model parameters at the global scale was developed. We used data from a diverse set of 1787 small-to-medium sized catchments (10–10,000 km^2) and the simple conceptual HBV model to set up and test the scheme. Each catchment was calibrated against observed daily *Q*, after which 674 catchments with high calibration and validation scores, and thus presumably good-quality observed *Q* and forcing data, were selected to serve as donor catchments. The calibrated parameter sets for the donors were subsequently transferred to 0.5° grid cells with similar climatic and physiographic characteristics, resulting in parameter maps for HBV with global coverage. For each grid cell, we used the 10 most similar donor catchments, rather than the single most similar donor, and averaged the resulting simulated *Q*, which enhanced model performance. The 1113 catchments not used as donors were used to independently evaluate the scheme. The regionalized parameters outperformed spatially uniform (i.e., averaged calibrated) parameters for 79% of the evaluation catchments. Substantial improvements were evident for all major Köppen-Geiger climate types and even for evaluation catchments > 5000 km distant from the donors. The median improvement was about half of the performance increase achieved through calibration. HBV with regionalized parameters outperformed nine state-of-the-art macroscale models, suggesting these might also benefit from the new regionalization scheme. The produced HBV parameter maps including ancillary data are available via www.gloh2o.org.

1. Introduction

All hydrologic models can to some degree benefit from calibration to improve their *Q* simulations, due to (i) lack of process understanding, (ii) possibly overly simplistic process representations, (iii) the spatiotemporal discretization of highly heterogeneous rainfall-runoff processes, and (iv) the impossibility of measuring all required model parameters at the model application scale [*Beven*, 1989; *Blöschl and Sivapalan*, 1995; *Duan et al*., 2001, 2006; *McDonnell et al*., 2007; *Nasonova et al*., 2009; *Rosero et al*., 2011; *Minville et al*., 2014]. Since *Q* observations are unavailable for the majority of the Earth's land surface [*Sivapalan*, 2003; *Hannah et al*., 2011], hydrologic models often rely on regionalization approaches to transfer information from gauged (donor) to ungauged (receptor) catchments [see *He et al*., 2011; *Hrachowitz et al*., 2013; *Razavi and Coulibaly*, 2013; *Blöschl et al*., 2013; *Parajka et al*., 2013 for reviews].

Six regionalization approaches have been used most frequently. First, the earliest regionalization approach consisted of catchment-by-catchment calibration and subsequent construction of a regression model that related the calibrated model parameters to catchment characteristics [e.g., *Seibert*, 1999; *Yokoo et al*., 2001; *Young*, 2006]. However, this approach generally met with limited success due in large part to the loss of model parameter interaction and the problem of equifinality [e.g., *Kokkonen et al*., 2003; *Hundecha and Bárdossy*, 2004; *McIntyre et al*., 2005; *Wagener and Wheater*, 2006; *Oudin et al*., 2008; *Kim and Kaluarachchi*, 2008]. Second, another widely used approach is to simultaneously construct the regression model and perform the calibration [e.g., *Hundecha and Bárdossy*, 2004; *Samaniego et al*., 2010], although this approach

involves the nontrivial task of formulating a priori parsimonious yet effective parameter-predictor relationships. Third, in another common approach, calibrated parameter sets are transferred to nearby regions based only on geographic proximity [e.g., *Merz and Blöschl*, 2004; *Oudin et al*., 2008]. However, this approach requires a dense network of gauging stations, and geographic proximity cannot necessarily be equated to similarity in rainfall-runoff behavior, particularly in climatically, geologically, or topographically highly heterogeneous regions [*Vandewiele and Elias*, 1995; *Shu and Burn*, 2003; *Oudin et al*., 2008; *Reichl et al*., 2009; *Ali et al*., 2012]. Alternatively (and fourthly), it is possible to transfer calibrated parameter sets based on explicit consideration of climatic and/or physiographic similarity [e.g., *Kokkonen et al*., 2003; *McIntyre et al*., 2005], or fifth simultaneously calibrate multiple catchments with similar climatic and/or physiographic characteristics to obtain more generalizable parameter sets [e.g., *Fernandez et al*., 2000; *Parajka et al*., 2007; *Kim and Kaluarachchi*, 2008]. However, these last two approaches are complicated by the need for a priori selection of characteristics representing the rainfall-runoff behavior. Sixthly and finally, various approaches have recently used Q signatures (measures that quantify the hydrograph shape such as slope of the flow-duration curve and baseflow index) estimated using, for example, regression to condition model parameters [e.g., *Yadav et al*., 2007; *Zhang et al*., 2008; *Troy et al*., 2008; *Castiglioni et al*., 2010]. However, this Q-signature approach is affected by: (i) the poor quality of daily meteorologic data in many regions which, after calibration against the estimated Q signatures, might lead to unrealistic model parameters producing the right results for the wrong reasons; (ii) the inherent difficulty in estimating Q signatures for ungauged regions [*Beck et al*., 2015]; and (iii) the fact that measures related to Q signatures by themselves are generally less effective at conditioning model parameters than goodness-of-fit measures.

Numerous studies have applied these approaches and demonstrated their respective advantages and limitations (see the aforementioned reviews). However, most studies had a regional (subcontinental) focus, employed a relatively small number of catchments, and used a variety of hydrologic models, forcing data, regionalization approaches, objective functions, and evaluation methodologies, resulting in findings with questionable generalizability [*Razavi and Coulibaly*, 2013]. In general, approaches that transfer calibrated parameter sets according to a certain measure of climatic and/or physiographic similarity performed better than (or gave comparable results to) other approaches [*Kokkonen et al*., 2003; *McIntyre et al*., 2004, 2005; *Parajka et al*., 2005; *Oudin et al*., 2008; *Li et al*., 2009; *Reichl et al*., 2009; *Bao et al*., 2012; *Wallner et al*., 2013; *Singh et al*., 2014; *Sellami et al*., 2014; *Garambois et al*., 2015]. There have been, however, several studies that obtained better results using approaches that transferred calibrated parameter sets based on geographic proximity [*Oudin et al*., 2008; *Zhang and Chiew*, 2009; *Samuel et al*., 2011; *Patil and Stieglitz*, 2014; *Petheram et al*., 2012], although the first four of these studies were conducted in regions with relatively dense gauging networks which tends to favor the spatial proximity approach (France, southeastern Australia, Ontario, and the conterminous USA, respectively). *Kay et al*. [2006] compared three regionalization approaches for two hydrologic models but obtained effectively inconclusive results: similarity-based regionalization performed best when using one hydrologic model, but performance was poorest in combination with the other model.

Due to the lack of a commonly accepted approach for parameter regionalization, hydrologic models typically applied at continental to global scales (hereafter called macroscale) rarely use regionalized parameters [*Sooda and Smakhtin*, 2015; *Kauffeldt*, 2014; *Bierkens et al*., 2015]. Instead, they tend to rely on a priori parameterizations based on expert opinion, case studies, field data, hydrologic theory, or data sets of questionable quality. For example, the Community Land Model (CLM) [*Oleson et al*., 2010], among many other models, uses a fixed value for the baseflow recession constant (k), although it has long been recognized that k varies spatially [*Hall*, 1968; *Beck et al*., 2013b]. Although the PCRaster Global Water Balance (PCR-GLOBWB) model [*Van Beek and Bierkens*, 2009] determines k based on drainage theory and hydrogeologic data, observational studies have reported weak links between k and current hydrogeologic data sets [*Van Dijk*, 2010; *Peña-Arancibia et al*., 2010; *Beck et al*., 2013b]. Similarly, several models (e.g., the Noah land surface model with Multi-Parameterization options—Noah-MP) [*Niu et al*., 2011] have adopted concepts from the TOPography based hydrologic MODEL (TOPMODEL) [*Beven and Kirkby*, 1979] to simulate surface runoff and baseflow, thus confounding model performance in regions where surface topography represents a poor proxy of the aquifer flow gradient and/or where other variables exert stronger controls on the flow response [*Beven*, 1997; *Devito et al*., 2005; *Li et al*., 2011]. Furthermore, in many models, including LISFLOOD [*Burek et al*., 2013], the generation of infiltration and saturation-excess runoff is implicitly linked to soil hydrologic

properties that are impossible to measure at the model application scale [*Blöschl and Sivapalan*, 1995; *Hopmans et al*., 2002]. Consequently, it is unlikely that current macro-scale hydrologic models have reached their full potential in terms of *Q* simulation [*Duan et al*., 2006; *Nasonova et al*., 2009; *Rosero et al*., 2011].

Table 1 lists (to the best of our knowledge) all macroscale attempts at model parameter regionalization, employing different catchment sets and descriptors, hydrologic models, regionalization approaches, and objective functions. However, none of these studies has explicitly quantified the improvement in *Q* simulation due to regionalization using a statistically significant number of independent catchments, nor did they compare the performance to a priori model parameters, rendering it impossible to verify the (added) value of the regionalization approaches used. In addition, two macroscale studies [*Widén-Nilsson et al*., 2007; *Troy et al*., 2008] used regionalization approaches based on spatial proximity or interpolation which should not be applied at macroscales, given that the majority of the land surface is ungauged or poorly gauged [*Sivapalan*, 2003; *Hannah et al*., 2011]. Moreover, three global studies [*Nijssen et al*., 2001; *Döll et al*., 2003; *Widén-Nilsson et al*., 2007] used only large catchments ($\gg$ 10,000 km^2), which typically tend to be more strongly affected by human activity (river regulation, diversions, water abstraction and extraction, and urbanization) and routing processes (evaporation from the river surface, riverbed leakage losses, and travel time delays) and thus are less likely to yield valid parameters at the grid-cell scale. Note that *Rakovec et al*. [2016] used large catchments as well, but explicitly accounted for the scale discrepancy problem.

Recently, using data from thousands of catchments around the globe, *Beck et al*. [2015] identified several climatic and physiographic characteristics that are strong predictors of *Q* signatures, ensuring that these characteristics are hydrologically relevant and that their data quality is sufficient. The hypothesis tested in the present study states that, at the global scale, similarity in these climatic and physiographic characteristics reflects (to a certain degree) similarity in rainfall-runoff response. We address this hypothesis by evaluating an easy-to-implement, global-scale regionalization scheme that overcomes the limitations of previous macroscale efforts. Our specific objectives are to: (i) identify catchments with good-quality observed *Q* and forcing data suitable to serve as donor catchments; (ii) produce parameter maps with global coverage by regionalizing the calibrated parameter sets of the donors; and (iii) explicitly quantify the improvement in model performance using regionalized parameters in independent catchments. To address these objectives, we use an unprecedentedly large, and highly diverse set of 1787 small-to-medium-sized catchments (10–10,000 km^2) around the globe in combination with the simple conceptual Hydrologiska Byråns Vattenbalansavdelning (HBV) hydrologic model [*Bergström*, 1992; *Seibert and Vis*, 2012].

2. Data and Methods

2.1. The HBV Hydrologic Model

Lumped and gridded versions of the HBV hydrologic model [*Bergström*, 1992; *Seibert and Vis*, 2012] were implemented in Python. HBV was chosen because of its flexibility, computational efficiency, proven effectiveness under a wide range of climatic and physiographic conditions [e.g., *Zhang and Lindström*, 1996; *Te Linde et al*., 2008; *Steele-Dunne et al*., 2008; *Breuer et al*., 2009; *Driessen et al*., 2010; *Deelstra et al*., 2010; *Plesca et al*., 2012; *Beck et al*., 2013a; *Bouffard*, 2014; *Demirel et al*., 2015; *Vetter et al*., 2015], and successful application in several previous regionalization studies [e.g., *Seibert*, 1999; *Hundecha and Bárdossy*, 2004; *Merz and Blöschl*, 2004; *Parajka et al*., 2005; *Booij*, 2005; *Parajka et al*., 2007; *Bárdossy*, 2007; *Jin et al*., 2009; *Masih et al*., 2010]. In addition, with its 14 calibratable parameters (of which five are related to the snow routine), HBV can arguably be considered to be of average complexity and thus fairly representative of a "typical" hydrologic model. HBV runs at a daily time step, has two groundwater stores and one unsaturated-zone store, and uses a triangular weighting function to simulate channel routing delays. The model requires daily time series of precipitation, potential evaporation, and air temperature as inputs. For each grid cell or catchment, only a single "elevation-vegetation zone" was considered here. Table 2 describes the model parameters. The stores were initialized by running the model for the first 10 year of the record if the record length was $\geq$ 10 year, or by running the model twice for the entire record if the record length was $<$10 year. The model was calibrated for each catchment for the time period with simultaneous observed *Q* and input data in a lumped fashion to reduce computational time. Table 2 lists the calibration ranges for each parameter. For more details concerning HBV, see *Bergström* [1992] and *Seibert and Vis* [2012].

Table 1. Studies Performing Model Parameter Regionalization at Macro (Continental to Global) Scales[a]

Study	Region	Model	Number of Regionalized Parameters	Regionalization Approach	Class	Number of Catchments for Parameterization/ Evaluation (size range)	Main Finding With Respect to Regionalization
Nijssen et al. [2001]	Global	Variable Infiltration Capacity (VIC)	6	Transfer of calibrated parameter sets to grid cells with the same climate	iv	9/13 (118,000 to 4,619,000 km^2)	Better performance was obtained for 6 of the 13 independent evaluation catchments
Döll et al. [2003]	Global	WaterGAP Global Hydrology Model (WGHM)	1	Regression model linking the model's "runoff coefficient" parameter to climatic, topographic, and morphologic variables	i	311/9 (>20,000 km^2)	The training coefficient of determination (R^2) for the regression model was 0.53
Boughton and Chiew [2007]	Australia	Australian Water Balance Model (AWBM)	6	Calibration against map of mean annual runoff produced using regional regression models based on mean annual precipitation and potential evaporation	vi	213/0 (50–2000 km^2)	The regression model training R^2 values ranged from 0.55 to 0.90
Widén-Nilsson et al. [2007]	Near global	Water and Snow Balance Modeling System Macroscale (WASMOD-M)	5	Transfer of calibrated parameter sets to grid cells within a 8.5° latitude by 19.5° longitude window	iii	485/0 (sizes not reported but appear to be ≫10,000 km^2)	Regionalized parameters produced good Q estimates, in contrast to spatially uniform parameters
Troy et al. [2008]	Conterminous USA	VIC	8	Calibration against map of runoff coefficient based on interpolation of observations	vi	1130/0 (10–10,000 km^2)	Resampling the regionalized parameters using averaging to a coarser grid led to markedly different Q estimates
Van Dijk et al. [2013]	Global	World-Wide Water Resources Assessment (W3RA)	1	Regression model linking the baseflow recession parameter to the aridity index [*Peña-Arancibia et al.*, 2010]	i	167/0 (<10,000 km^2)	The regression model training R^2 was 0.49
Livneh and Lettenmaier [2013]	Conterminous USA	Unified Land Model (ULM)	13	Regression model linking "zonally representative" parameters to catchment descriptors	i	220/0 (<10,000 km^2)	The regionalization performance deteriorated only slightly using catchment descriptors available globally
Singh et al. [2014]	Conterminous USA	Unnamed	8	Classification and regression tree (CART) analysis was used to optimize the similarity criterion	iv	83/0 (67–8151 km^2)	Climate and elevation were the most important for successful parameter regionalization
Bock et al. [2015]	Conterminous USA	Monthly Water Balance Model (MWBM)	6	Simultaneous calibration of catchments for 110 "hydrologically similar" regions	v	1575/0 (areas not specified but probably mostly <10,000 km^2)	Measured and simulated runoff showed good correspondence for the majority of the study region
Rakovec et al. [2016]	Europe	mesoscale Hydrologic Model (mHM)	28	Regression models linking 28 model parameters to 19 predictors [*Samaniego et al.*, 2010]	ii	400/0 (100–1,000,000 km^2)	The model was able to capture the Q dynamics well
This study	Global	Hydrologiska Byråns Vattenbalansavdelning (HBV)	13	Transfer of calibrated parameter sets to grid cells with similar climatic and physiographic characteristics	ii	674/1113 (<10,000 km^2)	Performance improvement due to regionalization for 79% of the independent evaluation catchments

[a]The studies are listed in order of publication date. The present study has been added for the sake of completeness. The number of evaluation catchments refers to independent catchments not used in any way for the parameterization. Regionalization approach classes are defined as: (i), catchment-by-catchment calibration followed by regression; (ii), simultaneous calibration and regression; (iii), geographic proximity; (iv), physiographic and/or climatic similarity; (v), regional calibration; and (vi), Q signatures.

2.2. Meteorological Forcing Data

For the catchment-scale calibration of HBV (described in section 2.4) and the performance evaluation of HBV using various parameter sets (described in section 2.6), we used the 0.5° Climate Prediction Center (CPC) Unified v1.0 precipitation data set (1979–2005) [*Xie et al.*, 2007; *Chen et al.*, 2008], which is based on

Table 2. HBV Model Parameter Descriptions and Calibration Ranges[a]

Parameter	Description	Minimum	Maximum
TT (°C)	Threshold temperature when precipitation is simulated as snowfall	−2.5	2.5
SFCF	Snowfall gauge undercatch correction factor	1	1.5
CWH	Water holding capacity of snow	0	0.2
CFMAX (mm $°C^{-1}$ d^{-1})	Melt rate of the snowpack	0.5	5
CFR	Refreezing coefficient	0	0.1
FC (mm)	Maximum water storage in the unsaturated-zone store	50	700
LP	Soil moisture value above which actual evaporation reaches potential evaporation	0.3	1
BETA	Shape coefficient of recharge function	1	6
UZL (mm)	Threshold parameter for extra outflow from upper zone	0	100
PERC (mm d^{-1})	Maximum percolation to lower zone	0	6
K0 (d^{-1})	Additional recession coefficient of upper groundwater store	0.05	0.99
K1 (d^{-1})	Recession coefficient of upper groundwater store	0.01	0.8
K2 (d^{-1})	Recession coefficient of lower groundwater store	0.001	0.15
MAXBAS (d)	Length of equilateral triangular weighting function	1	3

[a]The MAXBAS parameter was calibrated but not regionalized.

interpolation of gauge observations. We considered using the 0.5° WATCH Forcing Data ERA-Interim (WFDEI) meteorological data set (1979–2012) [*Weedon et al*., 2014], which is based on atmospheric reanalysis model output, but ended up using the CPC precipitation data set since it produced higher calibration scores for 70% of the catchments. This was not entirely unexpected, because gauges generally tend to provide the most accurate precipitation estimates in regions where the gauge density is sufficiently high [e.g., *Stillman et al*., 2016]. Conversely, for the performance comparison of HBV with regionalized parameters versus various state-of-the-art macro-scale models (also described in section 2.6), HBV was driven by precipitation from the WFDEI data set to be consistent with these other models.

The data sets for daily air temperature and potential evaporation remained the same throughout the study. Air temperature was derived from the WFDEI data set. Potential evaporation was calculated using the *Penman* [1948] formulation as given by *Shuttleworth* [1993], which was found to perform best in a comparison among five potential evaporation formulations [*Donohue et al*., 2010]. For the calculation of potential evaporation, daily net radiation, air temperature, atmospheric pressure, wind speed, and relative humidity were derived from the WFDEI data set, and surface albedo from a monthly climatology based on the European Space Agency (ESA) GlobAlbedo product [*Muller et al*., 2011].

2.3. Observed Streamflow Data

The observed daily *Q* and associated catchment boundary data were obtained from the same three sources as those used by *Beck et al*. [2015], namely the Geospatial Attributes of Gages for Evaluating Streamflow (GAGES)-II database [*Falcone et al*., 2010], the Global Runoff Data Centre (GRDC; http://www.bafg.de/GRDC/), and *Peel et al*. [2000]. The following eight criteria were used to exclude unsuitable catchments from our analysis:

1. The *Q* record length was required to be $\geq$ 10 y (not necessarily consecutive) during 1980–2005 (the common temporal span of the forcing used for the calibration of HBV).
2. To minimize the effects of channel routing, the catchment area had to be $<$10,000 km^2 (cf. *Lohmann et al*., 2004; *Peña-Arancibia et al*., 2010; *Xia et al*., 2012; *Van Dijk et al*., 2013; *Livneh and Lettenmaier*, 2013; *Beck et al*., 2015].
3. Since the climatic and physiographic data are likely less reliable at smaller scales, the minimum catchment area was taken as 10 km^2.
4. To reduce anthropogenic influences, catchments were required to have $<$ 2% (in total) classified as urban (using the "artificial areas" class of the GlobCover v2.3 map) [*Bontemps et al*., 2011] and subject to irrigation (using the Global Map of Irrigation Areas—GMIA; v4.0.1) [*Siebert et al*., 2005].
5. We used the Global Reservoir and Dam (GRanD) database (v1.1) [*Lehner et al*., 2011] to exclude catchments influenced by reservoirs (defined by total reservoir capacity $>$ 10% of the mean annual *Q*).

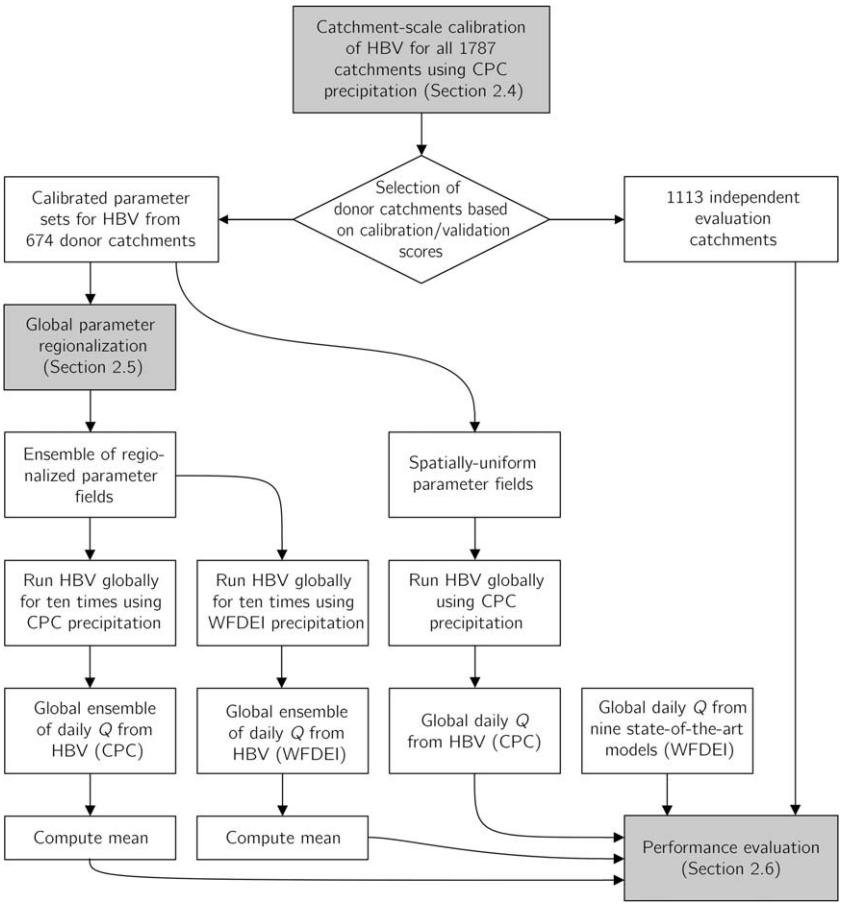


Figure 1. Flow chart summarizing the various steps carried out in this study to implement and test the regionalization scheme.

6. Catchments with forest gain or loss > 20% of the catchment area (the threshold at which changes in Q can be detected) [*Bosch and Hewlett*, 1982] were excluded using the Landsat-based forest change data set (v1.1) [*Hansen et al.*, 2013]. Although this data set only covers the period 2000–2013, currently no other reliable, high-resolution, global-scale data set on land-cover change exists.
7. Catchments in which water balance closure is impossible were discarded. These catchments were identified by $(Q_{obs}+PET_{Pen}) < P_{CPC}$ or $Q_{obs} > P_{CPC}$, where Q_{obs} is the mean annual observed Q (mm yr^{-1}), PET_{Pen} is the mean annual *Penman* [1948] potential evaporation (mm yr^{-1}), and P_{CPC} is the mean annual CPC precipitation (mm yr^{-1}) for the period 1980–2005 (see section 2.2).
8. To minimize the number of potentially disinformative catchments, all Q records were screened for artifacts and anthropogenic influences (diversions and/or impoundments), USA catchments flagged as "non-reference" in the GAGES-II database were discarded, and GRDC catchments for which the catchment boundaries could not be reliably determined were discarded [*Lehner*, 2012].

In total 1787 catchments (median size 505 km^2) fulfilled the selection criteria, the locations of which are presented in the Results section. All Q data were converted to mm d^{-1} using the provided catchment areas.

2.4. Catchment-Scale Calibration

Figure 1 presents a flow chart summarizing the various steps carried out to implement and test the regionalization scheme. The initial step involved calibration of the lumped version of HBV against observed daily Q for the selected 1787 catchments, first, to obtain calibrated parameter sets for the regionalization scheme and, second, to discard disinformative catchments with poor-quality observed Q and/or forcing data. For each catchment, the record of simultaneous forcing and observed Q data was split into a validation period (consisting of the first 30% of the record) and a calibration period (consisting of the remaining 70% of the

Table 3. The Q Signatures Incorporated in the AOF Objective Function[a]

Q Signature	Description and Computation	Transformation	Standard Deviation (σ)
BFI	Baseflow index, computed following the procedure described in *Institute of Hydrology* [1980] and *Gustard et al.* [1992].	None	0.12
Q1, Q10, Q50, Q90, Q99 (mm d^{-1})	Daily flow percentiles (exceedance probability). The number refers to the percentage of time that the flow is exceeded.	Square root	0.71, 0.49, 0.32, 0.19, 0.14
T50	The day of the water year marking the timing of the center of mass of *Q* [*Stewart et al.*, 2005]. The water year is defined in this study as October to September in the Northern Hemisphere and April to March in the Southern Hemisphere.	None	12.08
QMEAN (mm yr^{-1})	Mean annual *Q*.	Square root	7.41

[a]For Q1–99 and QMEAN, the standard deviation (σ) values were based on transformed values of the signatures.

record). For the calibration, we used an aggregate objective function (AOF) that considers both *Q* signatures and goodness-of-fit measures as suggested by *Vis et al.* [2015] and *Shafii and Tolson* [2015], rather than the commonly used *Nash and Sutcliffe* [1970] efficiency (NSE) which is widely considered to be a weak metric for model evaluation [e.g., *Schaefli and Gupta*, 2007; *Jain and Sudheer*, 2008; *Criss and Winston*, 2008; *Gupta et al.*, 2009]. The AOF score is computed following:

$$\mathrm{AOF}=\frac{\mathrm{AOF}_{\mathrm{sig}}+\mathrm{AOF}_{\mathrm{gof}}}{2}, \tag{1}$$

where $\mathrm{AOF}_{\mathrm{sig}}$ considers the signatures and $\mathrm{AOF}_{\mathrm{gof}}$ the goodness-of-fit of the simulated *Q* (all unitless). $\mathrm{AOF}_{\mathrm{sig}}$ incorporates the eight *Q* signatures listed in Table 3 and was defined by:

$$\mathrm{AOF}_{\mathrm{sig}}=1-\sum_{q=1}^{8}\frac{|Y_{q,o}-Y_{q,s}|}{8\sigma_q}, \tag{2}$$

where *Y* represent the transformed values of the signatures (unitless), σ the standard deviations of the transformed signatures (unitless), the *q* subscript denotes the *Q* signature, while the *o* and *s* subscripts refer to observed and simulated, respectively. Some *Q* signatures (Q1–99 and QMEAN) were square-root transformed prior to their inclusion in equation (2) to give small values more weight. The σ values in equation (2) represent the spatial variability in the *Q* signatures and serve to equalize the data variability (i.e., to give each transformed *Q* signature equal weight). They are listed in Table 3 and were derived from the Global Streamflow Characteristics Data set (GSCD) v1.9 [*Beck et al.*, 2015] taking into account the entire ice-free land surface excluding deserts (defined by an aridity index > 5), with the exception of the T50 σ, which considers only the snow-dominated ice-free land surface.

$\mathrm{AOF}_{\mathrm{gof}}$ incorporates three traditional goodness-of-fit measures and was computed following:

$$\mathrm{AOF}_{\mathrm{gof}}=\frac{2B+R+R_{\log}}{4}, \tag{3}$$

where *B* and *R* represent, respectively, the bias and Pearson correlation coefficient computed between simulated and observed *Q*, and $R_{\log}$ the Pearson correlation coefficient computed between natural-log transformed simulated and observed *Q* (all unitless). *B* was defined according to:

$$B=1-\left|\frac{\overline{Q_s}-\overline{Q_o}}{\overline{Q_s}+\overline{Q_o}}\right|, \tag{4}$$

where *Q* is the streamflow (mm d^{-1}), *s* and *o* as previously defined, and the overbars denote long-term averages. *B*, *R*, and $R_{\log}$ evaluate, respectively, the long-term flow volume, peak-flow variability, and low-flow variability. *B* ranges from 0 to 1, while *R* and $R_{\log}$ range from -1 to 1. To ensure that all three terms contribute equally to $\mathrm{AOF}_{\mathrm{gof}}$, *B* was multiplied by two in equation (3).

From equation (1) it follows that a higher AOF score corresponds to better model performance. We considered model simulations to be "unsatisfactory" if AOF ≤ 0.50, "satisfactory" if $0.50 <$ AOF ≤ 0.65, "good" if

$0.65 < \text{AOF} \leq 0.75$, and "very good" if $\text{AOF} > 0.75$. Accordingly, catchments with calibration and validation $\text{AOF} > 0.75$ were presumed to have good-quality observed Q and forcing data and deemed suitable to serve as donor catchments for the regionalization scheme [cf. *Bárdossy*, 2007; *Oudin et al*., 2008; *Bourgin et al*., 2015].

Various evolutionary algorithms have been applied in the calibration of hydrologic models [*Duan et al*., 1992; *Wang*, 1997; *Bekele and Nicklow*, 2007; *Maier et al*., 2014]. For the calibration of HBV, we used the $(\mu+\lambda)$ evolutionary algorithm implemented using the Distributed Evolutionary Algorithms in Python (DEAP) toolkit [*Fortin et al*., 2012]. The population size (μ) was set at 24 and the recombination pool size (λ) at 48. Each generation produced λ offspring from the population. Offspring were evaluated after which the population of the next generation was selected from both offspring and population. Crossover and mutation probabilities were set at 0.9 and 0.1, respectively. The number of generations was limited to 25, as this was found to be sufficient for achieving convergence. This amounted to 1200 model runs and AOF evaluations per catchment. The calibration of all 1787 catchments took approximately 35 h on a workstation with two Intel Xeon E5-2640 CPUs (total 16 cores and 32 threads).

2.5. Global-Scale Parameter Regionalization

After the catchment-scale calibration, we produced parameter maps (0.5° resolution) for HBV covering the entire ice-free land surface using a similarity-based regionalization approach that takes, for each grid cell, the calibrated parameter sets of the 10 most similar donor catchments (Figure 1). The 10 Q time series originating from the ten parameter sets were subsequently averaged to yield one single Q time series. A similarity-based approach was used since several previous studies have found similarity-based approaches to outperform other approaches [*Kokkonen et al*., 2003; *McIntyre et al*., 2004, 2005; *Parajka et al*., 2005; *Oudin et al*., 2008; *Li et al*., 2009; *Reichl et al*., 2009; *Bao et al*., 2012; *Wallner et al*., 2013; *Singh et al*., 2014; *Sellami et al*., 2014; *Garambois et al*., 2015]. The use of multiple donors ensures that the results are not dominated by individual donors with potentially unusual response behavior or unidentified data issues and has been found to enhance the model performance in several regionalization studies [*McIntyre et al*., 2004, 2005; *Oudin et al*., 2008; *Viney et al*., 2009; *Zhang and Chiew*, 2009; *Reichl et al*., 2009; *Bao et al*., 2012; *Zhang et al*., 2015; *Garambois et al*., 2015]. Ten donors were used since several of the aforementioned studies explicitly examined the optimal number of donors and achieved good results using ten donors [*McIntyre et al*., 2005; *Oudin et al*., 2008; *Zhang and Chiew*, 2009; *Reichl et al*., 2009; *Bao et al*., 2012]. Benefits of the proposed approach include: (i) its relative ease of implementation; (ii) retainment of model parameter interaction because the entire parameter set is transferred; (iii) possibility of spatial variability in model parameters according to landscape characteristics, even in ungauged regions; (iv) derived parameters are (largely) forcing independent; and (v) the use of multiple donor catchments (in this case 10) enables the estimation of parameter uncertainty.

The success of a similarity-based regionalization approach depends on the use of a large, highly diverse set of catchments (see section 2.3), high-quality observed Q and forcing data and therefore (more) reliable parameter estimates (see section 2.4), and a similarity criterion that represents the rainfall-runoff behavior of the catchments well [*Merz and Blöschl*, 2005; *Wagener et al*., 2007; *Reichl et al*., 2009; *Oudin et al*., 2010]. We used an a priori defined similarity criterion incorporating the eight climatic and physiographic characteristics listed in Table 4. These characteristics have been found to exhibit strong links with Q signatures in a previous global study [*Beck et al*., 2015], ensuring that they are relevant and that their data quality is sufficient. The dissimilarity between a catchment and grid-cell pair was quantified following:

$$S_{i,j} = \sum_{p=1}^{7} \frac{|Z_{p,i} - Z_{p,j}|}{\text{IQR}_p}, \tag{5}$$

where S is the dissimilarity (–), Z are the values of the respective characteristics (units listed in Table 4), IQR is the interquartile range of the characteristic (values and units listed in Table 4), p denotes the characteristic, and i and j denote, respectively, the catchment and grid cell in question. The IQR values represent the spatial variability in the various characteristics and were based on the ice-free land surface excluding deserts (defined by an aridity index > 5; see Table 4). The division by IQR in equation (5) was necessary to equalize the data variability of the characteristics. From equation (5) it follows that a similar catchment and grid-cell pair yields an S value close to zero. Catchment-mean values of characteristics were derived from

Table 4. The Climatic and Physiographic Characteristics Selected to Quantify the Similarity Between Catchments and Grid Cells

Variable	Description	Calculation and Data Source	Resolution	Interquartile Range (IQR)
AI	Aridity index	Calculated as: AI=PET/*P*, where *P* is the mean annual precipitation and PET the mean annual potential evaporation. Values were truncated with an upper limit of 10 to avoid extremely high values. See *P* and PET for data sources.	1 km	0.88
P (mm yr^{-1})	Mean annual precipitation	WorldClim v1.4 (release 3) [*Hijmans et al*., 2005], Parameter-elevation relationships on Independent Slopes Model (PRISM; *Daly et al*., 1994] for the USA, and *Tait et al*. [2006] for New Zealand.	1 km	743 mm yr^{-1}
PET (mm yr^{-1})	Mean annual potential evaporation	Calculated from monthly values derived following the temperature-based approach of *Hargreaves et al*. [1985]. See TA for data source.	1 km	1054 mm yr^{-1}
TA (°C)	Mean air temperature	WorldClim, and PRISM for the USA.	1 km	26.49°C
*f*TC	Fraction of forest cover	Landsat-based forest cover for the year 2000 (v1.1) [*Hansen et al*., 2013].	30 m	0.45
*f*S	Fraction of snow cover	Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua/Terra snow cover monthly Level 3 Global Climate Modeling Grid (CMG) product (MYD10CM/MOD10CM) v5 [*Hall et al*., 2006], mean over 2001–2014.	0.05°	0.57
SLO (°)	Surface slope	CGIAR Consortium for Spatial Information (CSI) Shuttle Radar Topography Mission (SRTM) v4.1 [*Farr et al*., 2007] for latitudes <60°N, GTOPO30 (http://lta.cr.usgs.gov/GTOPO30) for latitudes >60°N.	90 m, 1 km	1.08°
CLAY (%)	Soil clay content	SoilGrids1 km [*Hengl et al*., 2014] April 2014 version, mean over all layers.	1 km	13.77%

the full-resolution data, while the global-scale gridded data (0.5° resolution) were derived using simple averaging with gaps filled by nearest-neighbor interpolation. The ice-free portion of the land surface was determined using the World Wildlife Fund (WWF) Terrestrial Ecoregions of the World (TEOW) map [*Olson et al*., 2001].

The use of 10 parameter sets for each grid cell results in an ensemble of *Q* simulations of which the spread provides a valuable indication of the parameter uncertainty. It should be stressed, however, that these uncertainty estimates require careful interpretation as they are subject to the same criticisms as the widely used Generalized Likelihood Uncertainty Estimation (GLUE) approach [*Beven and Binley*, 1992], in that they lack a formal statistical foundation and involve several subjective choices (notably the calibration score threshold, the choice of 10 most similar donors, and the similarity criterion) [*McIntyre et al*., 2005; *Montanari*, 2005; *Winsemius et al*., 2009].

2.6. Performance of Regionalized Parameters

The value of the regionalization scheme was assessed in four ways. First, using the catchments having calibration and/or validation scores ≤ 0.75 which were rejected as donor (hereafter called the evaluation catchments; Figure 1), we compared AOF scores obtained by HBV using various sets of parameters:

1. spatially uniform parameters;
2. regionalized parameters based on the single most similar donor catchment (see section 2.5);
3. ensemble of regionalized parameters based on the 10 most similar donors (see section 2.5);
4. calibrated parameters for the validation period (see section 2.4); and
5. calibrated parameters for the calibration period (see section 2.4).

Parameter sets 1–3 represent the ungauged situation, while parameter sets 4 and 5 represent the ideal situation where observed *Q* data are available for calibration. To produce the spatially uniform parameters, the simple averages of the respective calibrated parameters for all donor catchments were computed [cf. *Kokkonen et al*., 2003; *Parajka et al*., 2005; *Kim and Kaluarachchi*, 2008; *Jin et al*., 2009]. For parameter sets

1–2, we ran the gridded version of HBV globally and computed catchment-mean Q time series. For parameter set 3, we ran the gridded model 10 times globally using the ensemble of regionalized parameters, subsequently computed the ensemble-mean Q, and finally computed catchment-mean Q time series. For parameter sets 1–3, the entire period of simultaneous observed and simulated Q was considered when computing the AOF score. To avoid mismatches between observed and simulated Q peaks from confounding the AOF scores, we introduced some channel routing delay to the catchment-mean simulated Q time series by applying the triangular weighting function of HBV [*Bergström*, 1992; *Seibert and Vis*, 2012] with the MAXBAS parameter (see Table 2) set to the calibrated value. For parameter sets 4 and 5, we used the AOF scores obtained in the catchment-scale calibration for the validation and calibration periods, respectively.

The second way we assessed the value of the regionalization scheme was by comparing AOF scores obtained for the evaluation catchments by HBV (with regionalized parameters based on the 10 most similar donors and spatially uniform parameters) to those obtained by nine state-of-the-art macroscale hydrologic models including their ensemble mean. The models were run globally as part of the eartH2Observe project and their simulated Q data were downloaded from https://wci.earth2observe.eu. All models were driven by the WFDEI meteorological data set, but used different formulations to compute potential evaporation, used different data sets for nonmeteorological variables, and were run at various spatial and temporal resolutions, although all outputs were resampled to a common 0.5° spatial and daily temporal resolution. Some of the models were subjected to varying degrees of calibration. For more details concerning the models, see *Dutra* [2015].

Third, we quantified the performance of HBV (with regionalized parameters based on the 10 most similar donors and with spatially uniform parameters) and the nine state-of-the-art models including their ensemble mean in the evaluation catchments in terms of more traditional performance metrics. This was done in order to (i) examine whether the performance improvement due to regionalization is restricted to the performance metric used for the calibration and (ii) allow the model performance to be put in the context of previous studies. The performance metrics considered were NSE, Kling-Gupta efficiency (KGE) [*Kling et al*., 2012], and coefficient of determination (R^2), all computed between daily, 5 day, and monthly mean (untransformed and log-transformed) simulated and observed Q. In addition, we considered an alternative bias-related performance metric, computed following:

$$B' = 1 - \left| \frac{\overline{Q_s} - \overline{Q_o}}{\overline{Q_o}} \right|, \tag{6}$$

where B' is the bias (unitless) and the other terms have been previously defined. For each performance metric, a higher value corresponds to a better model performance. HBV was run using WFDEI precipitation for this purpose.

The fourth and last way we assessed the value of the regionalization scheme was using a completely independent global QMEAN map (~0.04° resolution) based on Q observations from 1651 large catchments (10,003–4,691,000 km^2) around the globe (version 1.2) [*Beck*, 2016]. The map can be considered an updated version of the one produced by the University of New Hampshire (UNH) [*Fekete et al*., 2002] and was produced based on the assumption that the mean annual volumetric Q difference between a station and its upstream neighbor(s) represents the QMEAN generated in the interstation region. Specifically, we compared the QMEAN map of *Beck* [2016] to QMEAN maps derived from HBV with regionalized parameters (based on the 10 most similar donors) and spatially uniform parameters. For this comparison, HBV was run using WFDEI precipitation. Since the map of *Beck* [2016] provides spatially uniform values for the interstation regions, the HBV-based values were first averaged for the interstation regions.

3. Results

3.1. Catchment-Scale Calibration

Figure 2 shows the minimum values of the calibration and validation AOF scores as obtained with HBV when forced with CPC precipitation data for the study catchments, revealing a high degree of clustering. Clusters of well-performing catchments were found along the Pacific Coast of the USA, in the eastern USA, southern Great Britain, eastern Brazil, and southern Australia, while clusters of poorly performing catchments were found in the Interior West of the USA, the American tropics, and west of Lake Malawi (Africa).

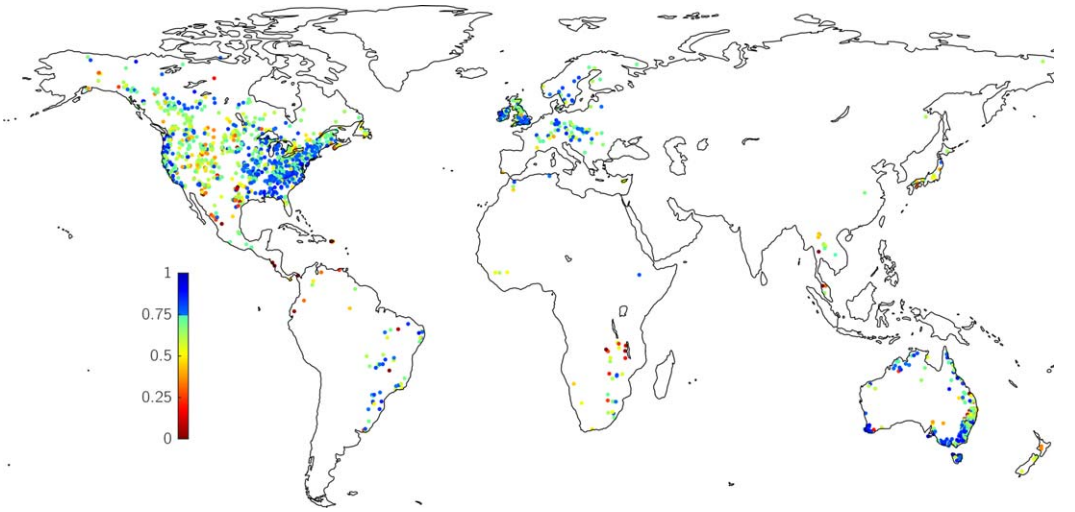


Figure 2. The minimum values of the calibration and validation AOF scores obtained using HBV for the study catchments. Each data point represents a catchment centroid ($n = 1787$). Catchments indicated in blue obtained calibration and validation scores > 0.75 and were used as donor catchment for the regionalization scheme. CPC precipitation was used to drive the model.

Table 5 lists the obtained median calibration and validation AOF scores for all catchments and each major Köppen-Geiger climate type (see supporting information section S1 for the newly derived world map of the Köppen-Geiger classification used here). The median calibration scores were very good for all climate types, ranging from 0.82 to 0.85. The decrease in median scores from calibration to validation ranged from 0.10 to 0.21, with the largest decreases found for tropical and arid catchments.

Among the 1787 study catchments, 674 had HBV-based calibration and validation AOF scores >0.75. These catchments presumably have good-quality meteorologic and observed *Q* data and thus were deemed suitable to serve as donors for the regionalization scheme. Table 5 also lists the number of donor catchments for each major Köppen-Geiger climate type. Temperate and cold climates were best represented by the donor catchments, tropical, and arid climates were moderately represented, while polar climates were poorly represented. The 1113 catchments not used as donor were used for independent evaluation of the improvement in model performance due to the regionalization. Table 5 shows that only the polar climate was poorly represented by the evaluation catchments.

3.2. Global-Scale Parameter Regionalization

The regionalization scheme transfers calibrated parameter sets from the donor catchments to similar grid cells to produce parameter maps covering the entire ice-free land surface. Figure 3a shows the spatial pattern for the mean dissimilarity to the 10 most similar donor catchments. The lowest mean dissimilarity was found for regions that are well represented by the donor catchments such as Europe and the USA, with especially low values obtained for the temperate eastern USA. Conversely, high mean dissimilarity was found for regions that are underrepresented by the donor catchments, such as tropical, polar, and mountainous regions (notably the Rocky Mountains, Andes, Himalayas, and Alps). Particularly high mean

Table 5. The Median Calibration and Validation AOF Scores and the Number of Donor and Evaluation Catchments[a]

Climate Type	Median Calibration AOF Score	Median Validation AOF Score	Number of Donor Catchments	Number of Evaluation Catchments
All	0.83	0.72	674	1113
A: tropical	0.83	0.63	15	61
B: arid	0.83	0.62	12	38
C: temperate	0.85	0.74	366	448
D: cold	0.82	0.71	277	560
E: polar	0.84	0.74	4	6

[a]The most dominant climate type by area was used to classify each catchment. See Figure 2 for the locations of the evaluation catchments and supporting information Figure S1.1 for the new world map of the Köppen-Geiger classification. HBV was forced with CPC precipitation.

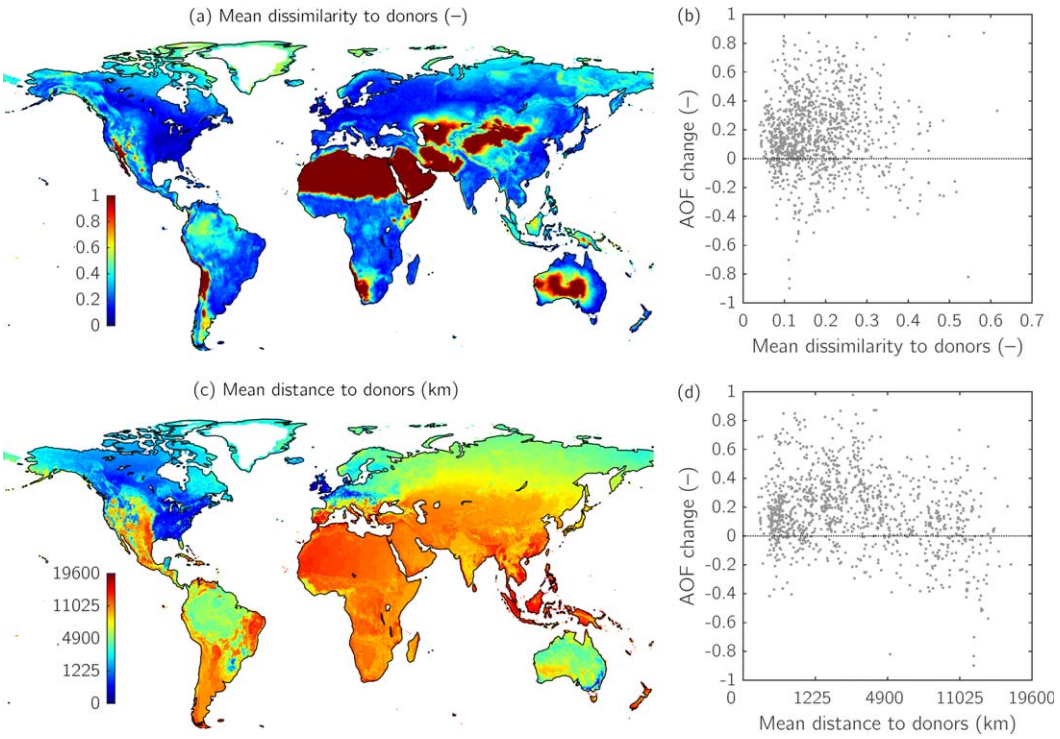


Figure 3. (a) Mean dissimilarity to the 10 most similar donor catchments. (b) Scatterplot of catchment-mean dissimilarity to the 10 most similar donor catchments versus AOF score obtained by HBV with regionalized parameters based on the 10 most similar donors minus the score obtained with spatially uniform parameters (AOF change; Figure 5). (c) Mean distance (as the "crow flies") to the 10 most similar donors (note the nonlinear color scale). (d) Scatterplot of catchment-mean distance to the 10 most similar donor catchments versus AOF change. Each data point in Figures 3b and 3d represents an evaluation catchment ($n = 1113$).

dissimilarity was found for arid regions. Figure 3c shows the mean distance (as the "crow flies") to the 10 most similar donor catchments. Mean distances of <1000 km were obtained only for densely gauged regions such as the USA, Europe, and southeastern Australia. Mean distances were generally >5000 km across South America, Africa, southern continental Asia, and Southeast Asia.

Figure 4 shows mean values of the regionalized HBV parameters based on the 10 most similar donor catchments. Only four key parameters are shown (see supporting information Figure S2.1 for maps of all other parameters). For the FC parameter (maximum water storage in the unsaturated-zone store), values were generally <300 mm everywhere except in tropical regions (Figure 4a). For the LP parameter (the soil moisture value above which actual evaporation reaches potential evaporation), generally values <0.65 were obtained for arid regions and values >0.80 for temperate, cold, and polar regions (Figure 4b). For the BETA parameter (shape coefficient of recharge function), values were consistently <2 for cold and polar regions and >3.5 for arid regions (Figure 4c). For the K2 parameter (recession coefficient of lower groundwater store), slow recessions ($K2 < 0.06$ d^{-1}) were typically found for cold and mountainous regions and fast recessions ($K2 > 0.10$ d^{-1}) for arid regions (Figure 4d). See supporting information Figure S2.2 for maps showing the standard deviation of the regionalized parameters.

3.3. Performance of Regionalized Parameters

Table 6 summarizes the performance in terms of median AOF score obtained in the 1113 evaluation catchments by HBV forced with CPC precipitation data using spatially uniform parameters, regionalized parameters (based on the single most similar and the 10 most similar donor catchments), and calibrated parameters (for the validation and calibration periods). Substantial improvements in median score ranging from 0.13 to 0.26 were obtained for the five Köppen-Geiger climate types using regionalized parameters based on the 10 most similar donors compared to spatially uniform parameters (Table 6). On the other hand, the improvements in median score using calibrated parameters for the validation period compared to spatially uniform parameters ranged from 0.31 to 0.38 (Table 6), reflecting the ideal situation where

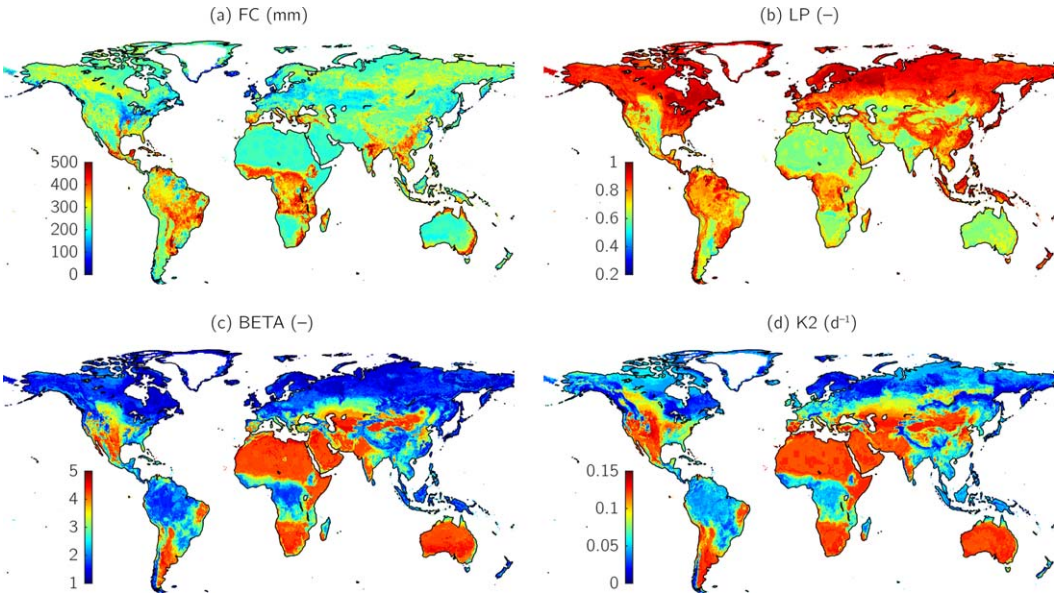


Figure 4. For HBV, mean values of the regionalized parameters based on the 10 most similar donor catchments for: (a) the maximum water storage in the unsaturated-zone store (FC); (b) the soil moisture value above which actual evaporation reaches potential evaporation (LP); (c) the shape coefficient of recharge function (BETA); and (d) the recession coefficient of lower groundwater store (K2). For maps of all other parameters, see supporting information Figure S2.1.

observed *Q* data are available for calibration. Thus, the performance improvement in terms of median AOF score obtained using regionalized parameters based on the 10 most similar donors was about half of that obtained using calibrated parameters for the validation period. For each grid cell, using the 10 most similar donors, rather than the single most similar donor, led in most cases to better model performance (Table 6). Figure 5 shows, for each evaluation catchment, the difference in AOF score using regionalized (based on the 10 most similar donors) versus spatially uniform parameters (called hereafter AOF change). For 79% of the evaluation catchments, the AOF change was positive, indicating better model performance with regionalized parameters (based on the 10 most similar donors) than with spatially uniform parameters.

Figure 3b presents a scatterplot for the evaluation catchments of catchment-mean dissimilarity to the 10 most similar donor catchments (Figure 3a) versus AOF change (Figure 5). No clear relationship can be discerned, suggesting that a lack of similar donor catchments does not necessarily diminish regionalization performance. Figure 3d shows the scatterplot of catchment-mean distance to the 10 most similar donor catchments (Figure 3c) versus AOF change (Figure 5), revealing that the greatest gains in performance were achieved for evaluation catchments situated $<$ 5000 km from the donors.

To further evaluate the effectiveness of the regionalization scheme, Table 7 compares median AOF scores obtained for the evaluation catchments by HBV with regionalized parameters (based on the 10 most similar

Table 6. The Performance in Terms of Median AOF Score Obtained for the Evaluation Catchments[a]

Climate Type	Spatially Uniform Parameters	Regionalized Parameters (Single Most Similar Donor)	Regionalized Parameters (10 Most Similar Donors)	Calibrated Parameters (Validation Period)	Calibrated Parameters (Calibration Period)
All (n = 1113)	0.30	0.46	0.49	0.64	0.79
A: tropical (n = 61)	0.21	0.21	0.34	0.57	0.81
B: arid (n = 38)	0.22	0.53	0.48	0.60	0.74
C: temperate (n = 448)	0.33	0.46	0.46	0.64	0.79
D: cold (n = 560)	0.29	0.48	0.52	0.65	0.79
E: polar (n = 6)	0.27	0.24	0.46	0.65	0.74

[a]See Figure 2 for the locations of the evaluation catchments and supporting information Figure S1.1 for the new world map of the Köppen-Geiger classification. HBV was forced with CPC precipitation.

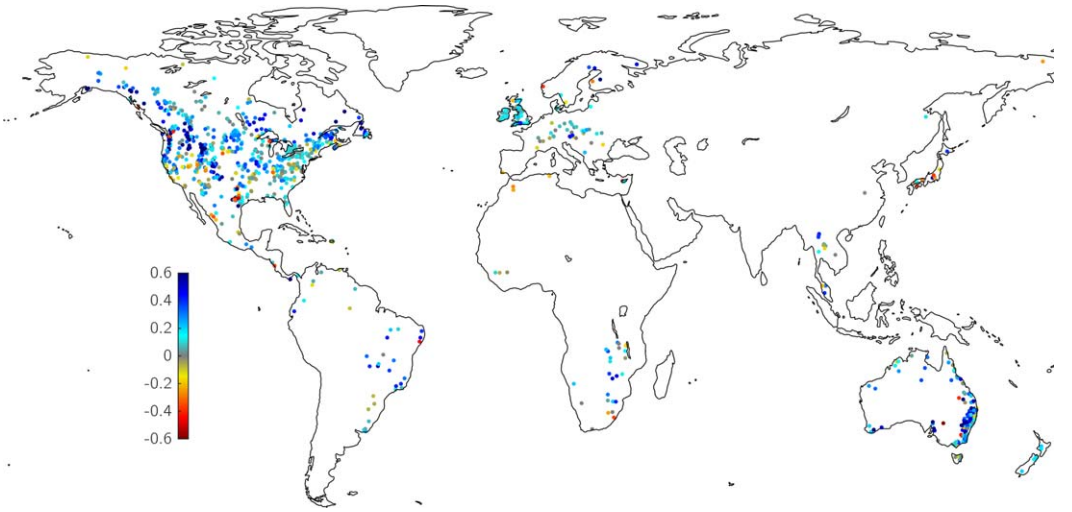


Figure 5. For the evaluation catchments, the AOF score obtained by HBV with regionalized parameters based on the 10 most similar donors minus the AOF score obtained using the spatially uniform parameters (defined as AOF change). Blue indicates enhanced model performance using regionalized parameters while yellow and red indicate diminished model performance using regionalized parameters. Each data point represents a catchment centroid ($n = 1113$). HBV was driven by CPC precipitation.

donors) and with spatially uniform parameters, to those obtained by nine state-of-the-art macroscale hydrologic models including their ensemble mean. For the evaluation catchments, the state-of-the-art models obtained median AOF scores ranging from 0.06 (SWBM) to 0.32 (WaterGAP3), while their ensemble mean gave a median AOF score of 0.35. HBV with spatially uniform parameters attained a median AOF score of 0.29, which is in the upper range of the median AOF scores obtained by the state-of-the-art models. Conversely, HBV with regionalized parameters based on the 10 most similar donors attained a much higher median AOF score of 0.47. The regionalized HBV model performed better than the other models for all climate types, although the ensemble mean obtained a slightly higher score for the tropics.

To examine whether the better performance translates to more widely used performance metrics and to allow the model performance to be put in the context of previous studies, Table 8 presents NSE, KGE, R^2, and B' scores obtained by HBV (using regionalized parameters derived from the 10 most similar donors and spatially uniform parameters) and the nine state-of-the-art models for the evaluation catchments. For all performance metrics, HBV with regionalized parameters performed either better or comparable to the other models, suggesting that the improved model performance due to the regionalization scheme is not restricted to the performance metrics used for calibration.

Figure 7 shows the difference in absolute square-root transformed QMEAN error between HBV with regionalized parameter (based on the 10 most similar donors) versus spatially uniform parameters, using the completely independent observation-based QMEAN map of *Beck* [2016] derived from catchments >10,000 km^2

Table 7. The Performance in Terms of Median AOF Score Obtained for the Evaluation Catchments ($n = 1113$)[a]

Climate Type	HBV With Regionalized Parameters (10 Most Similar Donors)	HBV With Spatially Uniform Parameters	HTESSEL	JULES	LISFLOOD	ORCHIDEE	PCR-GLOBWB	SURFEX	SWBM	W3RA	WaterGAP3	Ensemble Mean
All ($n = 1113$)	**0.47**	0.29	0.31	0.26	0.22	0.10	0.09	0.23	0.06	0.24	0.32	**0.35**
A: tropical ($n = 61$)	**0.36**	0.28	0.34	0.35	0.19	0.05	−0.47	0.22	0.09	0.17	0.32	**0.38**
B: arid ($n = 38$)	**0.52**	0.21	**0.43**	0.33	0.12	0.11	−0.20	0.33	0.36	0.41	0.33	0.32
C: temperate ($n = 448$)	**0.42**	0.31	0.31	0.36	0.07	0.16	0.01	**0.37**	0.18	0.23	0.30	0.33
D: cold ($n = 560$)	**0.51**	0.30	0.29	0.18	0.32	0.03	0.19	0.05	−0.05	0.23	0.34	**0.37**
E: polar ($n = 6$)	**0.54**	0.10	0.23	−0.10	**0.43**	0.15	0.30	−0.24	−0.31	0.04	0.27	0.39

[a]For clarity, in each row the two highest scores are shown in bold font. All models were forced with the WFDEI meteorological data set. The ensemble mean does not include HBV. See Figure 2 for the locations of the evaluation catchments, supporting information Figure S1.1 for the new world map of the Köppen-Geiger classification, and *Dutra* [2015] for descriptions of the state-of-the-art models.

Table 8. Median Scores for Widely Used Performance Metrics Obtained for the Evaluation Catchments ($n = 1113$)[a]

Performance Metrics	HBV With Regionalized Parameters (10 Most Similar Donors)	HBV With Spatially Uniform Parameters	HTESSEL	JULES	LISFLOOD	ORCHIDEE	PCR-GLOBWB	SURFEX	SWBM	W3RA	WaterGAP3	Ensemble Mean
NSE daily	**−0.02**	−0.03	−0.59	−0.38	−0.55	−0.45	−1.67	−0.24	**0.01**	−0.59	−0.11	−0.35
NSE 5 day	**0.08**	**0.05**	−0.49	−0.44	−0.26	−0.53	−1.51	−0.21	0.05	−0.34	−0.11	−0.25
NSE monthly	**0.17**	**0.15**	−0.32	−0.39	−0.03	−0.67	−1.16	−0.02	0.14	−0.10	−0.10	−0.09
NSE log-transformed daily	**−0.02**	−0.09	−0.46	−0.22	−0.69	−0.57	−1.25	**−0.06**	−0.08	−0.89	−0.40	−0.57
NSE log-transformed 5 day	**−0.00**	−0.09	−0.56	−0.32	−0.68	−0.88	−1.32	**−0.06**	−0.13	−0.90	−0.53	−0.62
NSE log-transformed monthly	**−0.05**	−0.14	−0.54	−0.41	−0.78	−1.45	−1.32	**−0.06**	−0.26	−0.96	−0.69	−0.69
KGE daily	**0.19**	0.11	−0.07	−0.03	−0.07	−0.18	−0.25	−0.08	0.11	−0.09	**0.13**	0.04
KGE 5 day	**0.26**	**0.20**	−0.02	0.02	0.11	−0.15	−0.19	−0.00	0.17	0.02	0.17	0.10
KGE monthly	**0.32**	**0.30**	0.12	0.12	0.29	−0.12	−0.07	0.09	0.26	0.18	0.23	0.22
KGE log-transformed daily	**0.16**	−0.13	0.08	0.04	−0.50	−0.10	−0.49	**0.12**	−0.07	−0.14	0.03	−0.05
KGE log-transformed 5 day	**0.15**	−0.12	0.11	0.09	−0.49	−0.06	−0.49	**0.14**	−0.05	−0.13	0.02	−0.06
KGE log-transformed monthly	**0.15**	−0.11	0.13	0.12	−0.51	−0.02	−0.49	**0.16**	−0.24	−0.12	−0.00	−0.06
R^2 daily	**0.36**	**0.30**	0.17	0.12	0.26	0.08	0.13	0.16	0.19	0.26	0.21	0.28
R^2 5 day	**0.45**	0.38	0.26	0.22	0.36	0.14	0.20	0.25	0.31	0.35	0.33	**0.40**
R^2 monthly	**0.57**	0.52	0.40	0.35	0.54	0.24	0.33	0.39	0.50	0.47	0.49	**0.55**
R^2 log-transformed daily	**0.50**	0.46	0.27	0.22	**0.48**	0.14	0.27	0.33	0.20	0.43	0.36	0.46
R^2 log-transformed 5 day	**0.54**	0.48	0.32	0.30	0.52	0.19	0.31	0.40	0.34	0.47	0.45	**0.54**
R^2 log-transformed monthly	**0.58**	0.52	0.40	0.42	0.56	0.27	0.38	0.48	0.51	0.53	0.55	**0.63**
$B\prime$	0.65	0.63	0.63	0.64	0.51	0.50	0.54	0.62	0.59	0.63	**0.66**	**0.65**

[a]For clarity, in each row the two highest scores are shown in bold font. All models were forced with the WFDEI meteorological data set. The ensemble mean does not include HBV. See Figure 1 of the main paper for the locations of the evaluation catchments, supporting information Figure S1.1 for the new world map of the Köppen-Geiger classification, and *Dutra* [2015] for descriptions of the state-of-the-art models.

as reference. HBV with regionalized parameters was found to perform better than HBV with spatially uniform parameters in terms of QMEAN for 68% of the gauged area, confirming the efficacy of the regionalization scheme. The QMEAN performance was clearly and consistently better for nearly all of Africa and Australia, while it was less in most of eastern Canada and China. The QMEAN performance was not clearly better or worse at high northern latitudes ($> 50°$N).

4. Discussion

4.1. Catchment-Scale Calibration

The patterns of HBV performance obtained for the USA (Figure 2) match those obtained in four previous studies using different hydrologic models and forcing data [*Lohmann et al*., 2004; *Xia et al*., 2012; *Newman et al*., 2015; *Bock et al*., 2015]. The consistently good model performance for the eastern USA (Figure 2) is probably due to the relatively homogeneous landscape and dense precipitation gauge network, whereas the poor model performance obtained for catchments in Central America, northern South America, and west of Lake Malawi (Africa; Figure 2) likely reflects the complex topography and lack of precipitation gauges in these areas [*Chen et al*., 2008]. Comparisons between observed and simulated Q data and between WorldClim and CPC precipitation data suggest that precipitation underestimation, caused by wind-induced undercatch as well as topographic bias in gauge placement [*Daly et al*., 1994; *Hijmans et al*., 2005; *Chen et al*., 2008], constitutes an important cause of poor model performance in humid catchments [cf. *Kauffeldt et al*., 2013]. The relatively large decrease in median calibration to validation scores obtained for tropical and arid catchments (Table 5) reflects the difficulty in estimating Q for these environments. The main confounding factors in the tropics are the frequent occurrence of short-duration, high-intensity convective storms, and the relatively low quality of the forcing and observed Q data [*Wohl et al*., 2012]. In arid regions, the main confounding factors are the high evaporative losses, the highly nonlinear response behavior, and the flashy nature of the Q [*Pilgrim et al*., 1988; *Ye et al*., 1997; *Lidén and Harlin*, 2000]. We suspect that model structural limitations exerted only a small influence on model performance, given the flexibility of HBV relative to other models [e.g., *Zhang and Lindström*, 1996; *Breuer et al*., 2009; *Deelstra et al*., 2010; *Plesca et al*., 2012; *Bouffard*, 2014; *Demirel et al*., 2015; *Vetter et al*., 2015].

4.2. Global-Scale Parameter Regionalization

The spatial patterns obtained for the regionalized HBV parameter values (Figure 4 and supporting information Figure S2.1) conform well with large-scale climate patterns, highlighting the dominant control of climate on the rainfall-runoff response. Most previous macroscale regionalization studies (Table 1) have used climate-related variables for the regionalization in one way or another, and thus recognized, either implicitly or explicitly, the link between model parameters and climate. In addition, several studies have shown that calibration to wet climatic periods leads to an overestimation of *Q* during dry climatic periods and vice versa [e.g., *Coron et al*., 2012; *Osuch et al*., 2015], suggesting that model parameters are related to climate not only in space but also in time. Conversely, regional regionalization studies usually omitted the use of climate-related variables [*He et al*., 2011], perhaps due to the relatively homogeneous climatic conditions prevailing in the study area, or because of the apparent misconception that models already account for climate through the meteorological forcing data [e.g., *Kokkonen et al*., 2003]. The use of climate-related variables to parameterize models is seemingly counterintuitive, given that models should represent only physiographic characteristics. However, climate is known to influence vegetation, soils, and geomorphology and thus exerts a major indirect influence on the rainfall-runoff behavior of catchments [*Gentine et al*., 2012; *Troch et al*., 2013]. Soil data are inherently uncertain due to (i) the scarcity and inconsistent quality and detail of soil information available around the world [*Hengl et al*., 2014], (ii) the difficulty in upscaling unevenly distributed point-scale soil profile data [*Hopmans et al*., 2002], (iii) the lack of information on soil macropore channels [*Beven and Germann*, 1982, 2013], and (iv) the knowledge gap with respect to the extent and severity of soil degradation around the globe [*Bai et al*., 2008]. This uncertainty undoubtedly translates to uncertainty in various model predictions which appears to be constrainable using climate-related variables. Furthermore, since models are by their very definition imperfect representations of reality, their state and flux estimates inevitably exhibit climate-dependent uncertainties, even if they are forced with "perfect" meteorological forcing data [*Beven*, 1989]. The tendency of hydrologic models to overestimate *Q* in arid regions [e.g., *Haddeland et al*., 2011; *Xia et al*., 2012; *Zhou et al*., 2012; *Trambauer et al*., 2013] and to generate the *Q* peak too early in snow-dominated regions [e.g., *Lohmann et al*., 2004; *Slater et al*., 2007; *Balsamo et al*., 2009; *Zaitchik et al*., 2010] can arguably be considered manifestations of these model imperfections.

The spatial pattern in mean regionalized FC (maximum water storage in the unsaturated-zone store) values (Figure 4a) appears to be somewhat random, suggesting that the parameter is less sensitive or subject to equifinality. Nevertheless, the slightly higher FC values in tropical regions could be indicative of the often relatively deep soils and regolith as well as the high water use and rainfall intercepting capacity of tropical rain forests [*Nepstad et al*., 1994; *Chappell et al*., 2007; *Tanaka et al*., 2008; *Holwerda et al*., 2012]. The low LP (soil moisture value above which actual evaporation reaches potential evaporation) and high BETA (shape coefficient of recharge function) parameter values obtained for arid regions (Figures 4b and 4c, respectively) serve to increase the evaporation and, conversely, decrease *Q*. The low LP value does so by increasing the rate of evaporation and the high BETA value by increasing the amount of rainfall assigned to the soil compartment. This suggest that, without regionalization, HBV would overestimate *Q* in arid regions, similar to many other hydrologic models [e.g., *Haddeland et al*., 2011; *Xia et al*., 2012; *Zhou et al*., 2012; *Trambauer et al*., 2013]. The K2 parameter (recession coefficient of lower groundwater store) map (Figure 4d) shows good consistency with observation-based maps for the pantropics [*Peña-Arancibia et al*., 2010] and the globe [*Beck et al*., 2013b, 2015], both in terms of spatial patterns and absolute magnitude. The fast recessions obtained for arid regions (Figure 4d) reflect the ephemeral nature of the quick flows that tend to dominate the flow regime under these conditions [*Pilgrim et al*., 1988].

4.3. Performance of Regionalized Parameters

Compared to previous macroscale regionalization studies (Table 1), we used a substantially more diverse set of 674 donor catchments with high calibration and validation scores and thus presumably (more) reliable calibrated parameters. Furthermore, we used the 10 most similar donor catchments for each grid cell, thereby providing a probabilistic estimate of *Q* of which the spread provides an indication of parameter uncertainty, and in addition used comparatively small catchments (10–10,000 km^2) to minimize the confounding influence of human activity and routing processes. Moreover, we explicitly quantified the performance of the scheme by comparing the performance of HBV with regionalized, spatially uniform, and calibrated parameters, and used an unprecedentedly large set of 1113 independent catchments to do so. This has likely led to more robust conclusions [cf. *Andréassian et al*., 2007; *Gupta et al*., 2014] and allowed us

to examine how climate type, donor similarity, and donor distance influence the performance of the regionalized HBV model.

HBV (forced with CPC precipitation) appears to perform markedly better with regionalized parameters than with spatially uniform parameters for the large majority of the 1113 independent evaluation catchments (Table 6 and Figure 5). Even for evaluation catchments located $>$ 5000 km away from the donors there were noticeable improvements in performance (Figure 3d). These findings confirm the value of the employed similarity criterion (equation (5)) and support the study hypothesis that, at the global scale, similarity in climate and physiography reflects (to a certain degree) similarity in rainfall-runoff response. The performance improvement for these more distant evaluation catchments was, however, slightly less (Figure 3d), suggesting there are still some functional differences among catchments unaccounted for. The omission of geology from the similarity criterion, due to its almost negligible Q predictive power [*Peña-Arancibia et al*., 2010; *Van Dijk*, 2010; *Beck et al*., 2013b,], potentially explains a small part of this unaccounted functional difference, while another part may be attributable to errors in the data sets used for the various climatic and physiographic characteristics (Table 4). The omission of geology means that the Q estimates for karst-dominated regions should be interpreted with caution. Interestingly, a lack of similar donor catchments did not noticeably influence model performance (Figure 3b), suggesting that the smaller number of donor catchments in tropical, arid, and polar climates (Table 5) was not detrimental to model performance. This is in contrast to *Sellami et al*. [2014], who found that the Q uncertainty increased as the similarity to donor catchments decreased, although they used only 10 catchments located in Mediterranean France.

The comparison between the HBV-based QMEAN maps versus the independent QMEAN map of *Beck* [2016] based on Q observations from catchments $>$ 10,000 km^2 further confirms the efficacy of the regionalization scheme and demonstrates that the improvement due to regionalization also translates to larger catchment scales (Figure 7). Particularly noteworthy is the good performance found for Africa, given the small number of donor catchments located in Africa. The mixed performance at high northern latitudes probably relates to the calibration of the snowfall gauge undercatch correction factor (SFCF) parameter, which yields forcing and location-dependent parameters more appropriate for catchment-scale applications than for regionalization. It should be noted that perfect agreement between the QMEAN map of *Beck* [2016] and the HBV-based QMEAN maps is unlikely, since the observed Q data used by *Beck* [2016] (i) are affected by water withdrawals [*Döll et al*., 2003] and transmission losses [*Lange*, 2005] which HBV does not account for and (ii) cover a different time period than the HBV-based simulated Q data.

Comparing the presently obtained improvements in HBV model performance due to the regionalization scheme to previous studies is not straightforward, because most other studies had a regional rather than global focus and typically used a smaller number of catchments and different hydrologic models and catchment variables, as well as different regionalization and evaluation approaches (see Table 1 and reviews by *He et al*. [2011]; *Hrachowitz et al*. [2013]; *Razavi and Coulibaly* [2013]; *Blöschl et al*. [2013]; *Parajka et al*. [2013], and references therein). Moreover, most previous studies used the NSE performance metric, were confined to humid settings (aridity index $<$ 1), and did not explicitly quantify the performance improvement due to regionalization. Nevertheless, in agreement with the present results (Table 6 and Figure 5), previous studies obtained good model performance using similarity-based regionalization for Austria [*Parajka et al*., 2005], southeastern Australia [*Li et al*., 2009; *Reichl et al*., 2009], China [*Bao et al*., 2012], France [*Oudin et al*., 2008], Mediterranean France [*Sellami et al*., 2014; *Garambois et al*., 2015], northern Germany [*Wallner et al*., 2013], North Carolina (USA) [*Kokkonen et al*., 2003], the UK [*McIntyre et al*., 2004, 2005], and the conterminous USA [*Singh et al*., 2014]. The substantially better performance achieved here using the 10 most similar donor catchments, rather than the single most similar donor (Table 6), reaffirms the importance of using parameter sets from multiple donors [cf. *McIntyre et al*., 2004, 2005; *Oudin et al*., 2008; *Viney et al*., 2009; *Zhang and Chiew*, 2009; *Reichl et al*., 2009; *Bao et al*., 2012; *Zhang et al*., 2015; *Garambois et al*., 2015]. *Parajka et al*. [2013] reported in their review that studies generally found poorer regionalization performance for smaller and more arid catchments. In the present analysis, however, the AOF score obtained by HBV with regionalized parameters was related to neither catchment area nor aridity index (Figure 6). Instead, HBV with regionalized parameters displayed a markedly lower median AOF score for the tropical catchments (Table 6).

The improved performance of HBV due to regionalization appears to translate also to performance metrics not explicitly incorporated in the objective function used for calibration (Table 8). However, both HBV and

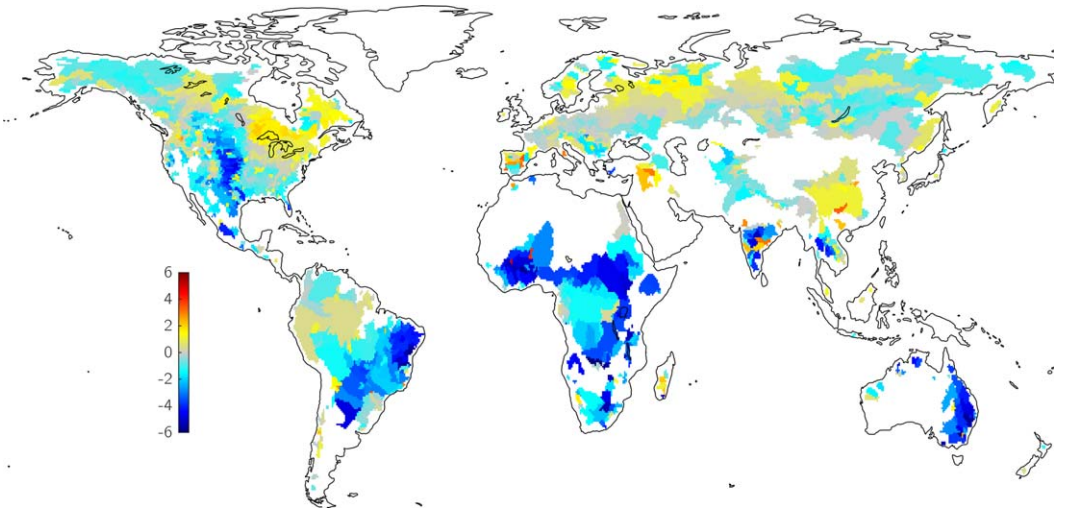


Figure 6. The difference in absolute square-root transformed QMEAN error between HBV with regionalized parameters (based on the 10 most similar donor catchments) versus spatially uniform parameters, using the observation-based QMEAN map of *Beck* [2016] as reference. Blue indicates HBV with regionalized parameters is closer to the observations, whereas yellow and red indicate HBV with spatially-uniform parameters is closer to the observations. White indicates that no observations were available. WFDEI precipitation was used to drive HBV.

the nine state-of-the-art macroscale models generally showed rather poor scores compared to previous studies, particularly in terms of NSE. For example, we obtained NSE values computed from daily (untransformed) Q data ranging from -1.67 (PCR-GLOBWB) to -0.02 (HBV with regionalized parameters). These values are of similar magnitude to the daily NSE values obtained by *Xia et al*. [2012] using five (uncalibrated) macroscale hydrologic models in 961 small-to-medium sized U.S. catchments (23–10,000 km^2). However, our values are considerably lower than the range in daily NSE values of 0.50–0.81 (median 0.66) found for nine previous similarity-based regionalization studies [*Parajka et al*., 2013]. The lower performance obtained in the present work probably reflects: (i) the use of evaluation catchments with low-quality observed Q and/or forcing data, due to the exclusion of catchments with calibration or validation AOF scores ≤ 0.75; (ii) the use of relatively coarse 0.5° forcing data; (iii) the tremendous climatic and physiographic diversity of the catchments included in the present study; and (iv) a generally (much) greater distance between donor and evaluation catchments.

HBV forced with WFDEI precipitation also performed considerably better with regionalized parameters than with spatially uniform parameters (Table 7), suggesting that the effectiveness of the regionalization scheme is not restricted to the forcing data used for calibration. Furthermore, HBV with regionalized parameters outperformed nine state-of-the-art macroscale models including their ensemble mean (Table 7), suggesting

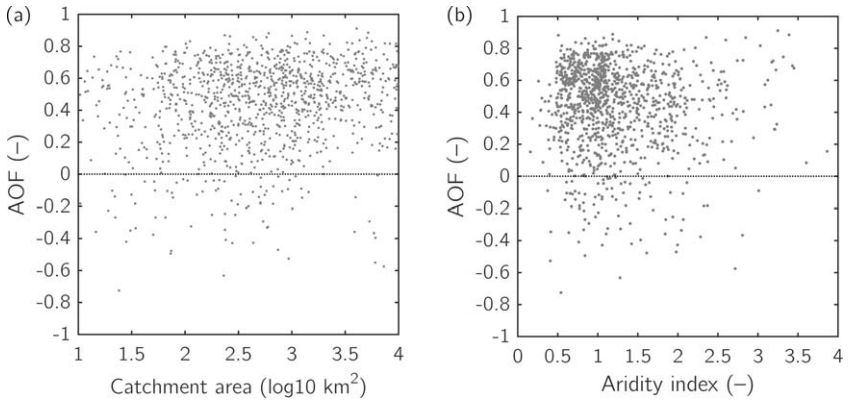


Figure 7. Scatterplots for the evaluation catchments of (a) catchment area and (b) catchment-mean aridity index versus the AOF score obtained by HBV with regionalized parameters derived from the 10 most similar donors. Each data point represents an evaluation catchment ($n = 1113$).

that these models can also benefit from application of the new regionalization scheme. However, the potential performance improvement will depend on the structure, parameterization, and forcing of the model in question. Many current models have an inflexible structure and use a priori parameters and thus cannot be calibrated successfully [*Mendoza et al*., 2015], although *Parajka et al*. [2013] indicated in their review that studies generally showed poorer regionalization performance with a higher number of calibratable model parameters. This suggests that one must strike the right balance between model flexibility and the number of calibratable parameters for optimal results. The quality of the forcing data further influences the potential improvement, by imposing limits on the maximum attainable performance [e.g., *Van Dijk et al*., 2013; *Kauffeldt et al*., 2013].

4.4. Future Work

Although the presently produced parameter maps constitute a clear improvement over the common practice of using uniform parameter values for the entire globe, there may be room for improvement with respect to, among other things, the objective function (equation (1)), the donor selection criterion, the similarity criterion (equation (5)), as well as the ensemble size. Expanding the number of donor catchments for tropical, arid, and polar climates may also be of interest for future studies. Non-*Q*-related aspects of the model may be improved by incorporating information on relevant hydrologic variables in the objective function, for example GRACE-based total water storage [e.g., *Eicker et al*., 2014], satellite-derived surface soil moisture [e.g., *Wanders et al*., 2014], and remotely sensed snow cover [e.g., *Duethmann et al*., 2014]. We welcome additional verification efforts using *Q* data from catchments not used in the present effort as well as global-scale comparisons against alternative regionalization approaches (notably the approach of *Samaniego et al*., 2010). The presently obtained results may be used as a possible performance baseline for future improvements. Besides these technical improvements, it is essential to improve understanding of rainfall-runoff processes under different physiographic and climatic conditions in an effort to improve the structure and parameterization of our models and consequently reduce the need for calibration and regionalization [cf. *Hrachowitz et al*., 2013; *Nijzink et al*., 2015].

5. Conclusions

The present study is the first to demonstrate improved *Q* simulation due to hydrologic model parameter regionalization at the global scale, providing support for the hypothesis that similarity in climate and physiography reflects (to a certain degree) similarity in rainfall-runoff behavior. The main conclusions reached are:

1. Precipitation underestimation appeared to be the dominant cause of low calibration AOF scores obtained for HBV. Relatively large decreases from calibration to validation scores were found for tropical and arid catchments. Among the 1787 investigated catchments, 674 achieved calibration and validation scores >0.75 and thus were deemed suitable to serve as donors for the regionalization scheme. Tropical, arid, and polar climates were somewhat underrepresented among the donors.
2. The regionalization scheme transfers calibrated parameter sets from the donor catchments to similar grid cells to produce parameter maps for HBV covering the entire ice-free land surface. The spatial patterns in regionalized parameter values corresponded well with spatial patterns in climate, which conflicts with the common practice of parameterizing hydrologic models based on physiographic properties only.
3. The 1113 catchments not used as donors were used to independently quantify the improvement in HBV-based *Q* estimates due to regionalization. The regionalized parameters based on the 10 most similar donors produced better *Q* estimates than did spatially uniform parameters for most (79%) of the evaluation catchments. Substantial improvements were achieved for all major Köppen-Geiger climate types and even evaluation catchments located >5000 km distance from the donors. The median improvement in performance was about half the increase that was achieved through calibration. HBV with regionalized parameters also outperformed nine state-of-the-art macroscale models including their ensemble mean, suggesting that these models could indeed benefit from application of the currently developed regionalization scheme.

Acknowledgments
The produced HBV parameter maps and ancillary data are available via www.gloh2o.org. The WorldClim developers are thanked for making available the global climate data and the SoilGrids1km developers for providing the global soil data. We gratefully acknowledge the Global Runoff Data Centre (GRDC) and the U.S. Geological Survey (USGS) for providing most of the observed *Q* data used in the present study. We particularly thank Wouter Berghuijs for his detailed and constructive comments, as well as Luis Samaniego and two anonymous reviewers for their useful criticism on earlier drafts, and suggestions for improvement. This research received partial funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grand agreement 603608, "Global earth observation for integrated water resource assessment": eartH2Observe. D.G.M. acknowledges financial support from The Netherlands Organization for Scientific Research through grant 863.14.004. The views expressed herein are those of the authors and do not necessarily reflect those of the European Commission.

References

Ali, G., D. Tetzlaff, C. Soulsby, J. J. McDonnell, and R. Capell (2012), A comparison of similarity indices for catchment classification using a cross-regional dataset, *Adv. Water Resour.*, *40*, 11–22, doi:10.1016/j.advwatres.2012.01. 008.

Andréassian, V., J. Lerat, C. Loumagne, T. Mathevet, C. Michel, L. Oudin, and C. Perrin (2007), What is really undermining hydrologic science today?, *Hydrol. Processes*, *21*(20), 2819–2822.

Bai, Z. G., D. L. Dent, L. Olsson, and M. E. Schaepman (2008), Proxy global assessment of land degradation, *Soil Use Manage.*, *24*(3), 223–234, doi:10.1111/j.1475-2743.2008.00169.x.
Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts (2009), A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system, *J. Hydrometeorol.*, *10*(3), 623–643.
Bao, Z., J. Zhang, J. Liu, G. Fu, G. Wang, R. He, X. Yan, J. Jin, and H. Liu (2012), Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-climatic conditions, *J. Hydrol.*, *466–467*(1), 37–46.
Bárdossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, *11*, 703–710.
Beck, H. E. (2016), A global map of mean annual runoff based on discharge observations from large catchments, doi:10.5281/zenodo.44782, in press.
Beck, H. E., L. A. Bruijnzeel, A. I. J. M. van Dijk, T. R. McVicar, F. N. Scatena, and J. Schellekens (2013a), The impact of forest regeneration on streamflow in 12 meso-scale humid tropical catchments, *Hydrol. Earth Syst. Sci.*, *17*(7), 2613–2635.
Beck, H. E., A. I. J. M. van Dijk, D. G. Miralles, R. A. M. de Jeu, L. A. Bruijnzeel, T. R. McVicar, and J. Schellekens (2013b), Global patterns in baseflow index and recession based on streamflow observations from 3394 catchments, *Water Resources Research*, *49*, 7843–7863, doi:10.1002/2013WR013918.
Beck, H. E., A. I. J. M. van Dijk, and A. de Roo (2015), Global maps of streamflow characteristics based on observations from several thousand catchments, *J. Hydrometeorol.*, *16*(4), 1478–1501.
Bekele, E. G., and J. W. Nicklow (2007), Multi-objective automatic calibration of SWAT using NSGA-II, *J. Hydrol.*, *341*(3–4), 165–176.
Bergström, S. (1992), The HBV model—its structure and applications, *SMHI Rep. RH 4*, Swed. Meteorol. and Hydrol. Inst., Norrköping, Swed.
Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298.
Beven, K., and P. Germann (1982), Macropores and water flow in soils, *Water Resour. Res.*, *18*(5), 1311–1325.
Beven, K., and P. Germann (2013), Macropores and water flow in soils revisited, *Water Resour. Res.*, *49*, 3071–3092, doi:10.1002/wrcr.20156.
Beven, K. J. (1989), Changing ideas in hydrology: The case of physically-based models, *J. Hydrol.*, *105*(1–2), 157–172.
Beven, K. J. (1997), TOPMODEL: A critique, *Hydrol. Processes*, *11*(9), 1069–1085.
Beven, K. J., and M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, *24*(1), 43–69.
Bierkens, M. F. P., et al. (2015), Hyper-resolution global hydrological modelling: What is next?, *Hydrol. Processes*, *29*(2), 310–320.
Blöschl, G., and M. Sivapalan (1995), Scale issues in hydrological modelling: A review, *Hydrol. Processes*, *9*(3–4), 251–290.
Blöschl, G., M. Sivapalan, T. Wagener, A. Viglione, and H. Savenije (Eds.) (2013), *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*, Cambridge Univ. Press, N. Y.
Bock, A. R., L. E. Hay, G. J. McCabe, S. L. Markstrom, and R. D. Atkinson (2015), Parameter regionalization of a monthly water balance model for the conterminous United States, *Hydrol. Earth Syst. Sci. Discuss.*, *12*(9), 10,023–10,066.
Bontemps, S., P. Defourny, E. Van Bogaert, O. Arino, V. Kalogirou, and J. J. Ramos Perez (2011), GlobCover 2009, products description and validation report, technical report, 53 pp., ESA GlobCover Proj. [Available at http://ionia1.esrin.esa.int.]
Booij, M. J. (2005), Impact of climate change on river flooding assessed with different spatial model resolutions, *J. Hydrol.*, *303*(1–4), 176–198.
Bosch, J. M., and J. D. Hewlett (1982), A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration, *J. Hydrol.*, *55*(1–4), 3–23.
Bouffard, J.-S. (2014), A comparison of conceptual rainfall-runoff modelling structures and approaches for hydrologic prediction in ungauged peatland basins of the James Bay lowlands, Master's thesis, Carleton Univ., Ottawa, Ont.
Boughton, W., and F. Chiew (2007), Estimating runoff in ungauged catchments from rainfall, PET and the AWBM model, *Environ. Modell. Software*, *22*(4), 476–487.
Bourgin, F., V. Andréassian, C. Perrin, and L. Oudin (2015), Transferring model uncertainty estimates from gauged to ungauged catchments, *Hydrol. Earth Syst. Sci.*, *19*(5), 2535–2546.
Breuer, L., et al. (2009), Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use, *Adv. Water Resour.*, *32*(2), 129–146.
Burek, P., J. van der Knijff, and A. de Roo (2013), LISFLOOD distributed water balance and flood simulation model revised user manual, *Tech. Rep. EUR 26162 EN*, Joint Res. Cent., Ispra, Italy, doi:10.2788/24719.
Castiglioni, S., L. Lombardi, E. Tot, A. Castellarin, and A. Montanari (2010), Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach, *Adv. Water Resour.*, *33*(10), 1235–1242.
Chappell, N. A., M. Sherlock, K. Bidin, R. Macdonald, Y. Najman, and G. Davies (2007), Runoff processes in Southeast Asia: Role of soil, regolith, and rock type, in *Forest Environments in the Mekong River Basin*, edited by H. Sawada, et al., Springer, Tokyo, Japan.
Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak (2008), Assessing objective techniques for gauge-based analyses of global daily precipitation, *J. Geophys. Res. 113*, D04110, doi:10.1029/2007JD009132.
Coron, L., V. Andréassian, C. Perrin, J. Lerat, J. Vaze, M. Bourqui, and F. Hendrickx (2012), Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, *Water Resour. Res.*, *4*, W05552, doi:10.1029/2011WR011721.
Criss, R. E., and W. E. Winston (2008), Do Nash values have value? Discussion and alternate proposals, *Hydrol. Processes*, *22*(14), 2723–2725.
Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*(2), 140–158.
Deelstra, J., C. Farkas, A. Engebretsen, S. Kværnø, S. Beldring, A. Olszewska, and L. Nesheim (2010), Can we simulate runoff from agriculture dominated watersheds? Comparison of the DrainMod, SWAT, HBV, COUP and INCA models applied for the Skuterud catchment, *Bioforsk FOKUS*, *5*(6), 119–128.
Demirel, M. C., M. J. Booij, and A. Y. Hoekstra (2015), The skill of seasonal ensemble low-flow forecasts in the Moselle River for three different hydrological models, *Hydrol. Earth Syst. Sci.*, *19*(1), 275–291.
Devito, K., I. Creed, T. Gan, C. Mendoza, R. Petrone, U. Silins, and B. Smerdon (2005), A framework for broad-scale classification of hydrologic response units on the Boreal Plain: Is topography the last thing to consider?, *Hydrol. Processes*, *19*(8), 1705–1714.
Döll, P., F. Kaspar, and B. Lehner (2003), A global hydrological model for deriving water availability indicators: Model tuning and validation, *J. Hydrol.*, *270*(1), 105–134.
Donohue, R. J., T. R. McVicar, and M. L. Roderick (2010), Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate, *J. Hydrol.*, *386*(1–4), 186–197.
Driessen, T. L. A., R. T. W. L. Hurkmans, W. Terink, P. Hazenberg, P. J. J. F. Torfs, and R. Uijlenhoet (2010), The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model, *Hydrol. Earth Syst. Sci.*, *14*(4), 651–665.
Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031.

Duan, Q., J. Schaake, and V. Koren (2001), *A Priori* estimation of land surface model parameters, in *Land Surface Hydrology, Meteorology, and Climate: Observations and Modeling, Water Sci. Appl*., vol. 3, edited by V. Lakshmi, J. Albertson, and J. Schaake, pp. 77–94, AGU, Washington, D. C.
Duan, Q., et al. (2006), Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *J. Hydrol.*, *320*(1), 3–17.
Duethmann, D., J. Peters, T. Blume, S. Vorogushyn, and A. Güntner (2014), The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia, *Water Resour. Res.*, *50*, 2002–2021, doi:10.1002/2013WR014382.
Dutra, E. (2015), Report on the current state-of-the-art Water Resources Reanalysis, *Tech. Rep. D.5.1*, EartH2Observe. [Available at http://earth2observe.eu/files/PublicDeliverables/D5.1_Report on the WRR1 tier1.pdf.]
Eicker, A., M. Schumacher, J. Kusche, P. Döll, and H. Müller Schmied (2014), Calibration/data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an ensemble Kalman filter: First results, *Surv. Geophys.*, *35*(6), 1285–1309.
Falcone, J. A., D. M. Carlisle, D. M. Wolock, and M. R. Meador (2010), GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States, *Ecology*, *91*(2), 621 pp.
Farr, T. G., et al. (2007), The shuttle radar topography mission, *Rev. Geophys.*, *45*, RG2004, doi:10.1029/2005RG000183.
Fekete, B. M., C. J. Vörösmarty, and W. Grabs (2002), High-resolution fields of global runoff combining observed river discharge and simulated water balances, *Global Biogeochem. Cycles*, *16*(3), 1042, doi:10.1029/1999GB001254.
Fernandez, W., R. M. Vogel, and A. Sankarasubramanian (2000), Regional calibration of a watershed model, *Hydrol. Sci. J.*, *45*(5), 689–707.
Fortin, F., F. De Rainville, M. Gardner, M. Parizeau, and C. Gagné (2012), DEAP: Evolutionary algorithms made easy, *J. Mach. Learn. Res.*, *13*, 2171–2175.
Garambois, P. A., H. Roux, K. Larnier, D. Labat, and D. Dartus (2015), Parameter regionalization for a process-oriented distributed model dedicated to flash floods, *J. Hydrol.*, *525*, 383–399, doi:10.1016/j.jhydrol.2015.03.052.
Gentine, P., P. D'Odorico, B. R. Litner, G. Sivandran, and G. Salvucci (2012), Interdependence of climate, soil, and vegetation as constrained by the Budyko curve, *Geophys. Res. Lett.*, *39*, L19404, doi:10.1029/2012GL053492.
Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, *370*(1–2), 80–91.
Gupta, H. V., C. Perrin, R. Kumar, G. Blöschl, M. Clark, A. Montanari, and V. Andréassian (2014), Large-sample hydrology: A need to balance depth with breadth, *Hydrol. Earth Syst. Sci.*, *18*, 463–477.
Gustard, A., A. Bullock, and J. M. Dixon (1992), Low flow estimation in the United Kingdom, *Tech. Rep. 108*, Inst. of Hydrol., Wallingford, U. K.
Haddeland, I., et al. (2011), Multimodel estimate of the global terrestrial water balance: Setup and first results, *J. Hydrometeorol.*, *12*(5), 869–884.
Hall, D. K., V. V. Salomonson, and G. A. Riggs (2006), MODIS/Aqua snow cover daily L3 global 0.05deg CMG. Version 5, technical report, Natl. Snow and Ice Data Cent., Boulder, Colo.
Hall, F. R. (1968), Base-flow recessions: A review, *Water Resour. Res.*, *4*(5), 973–983.
Hannah, D. M., S. Demuth, H. A. J. Van Lanen, U. Looser, C. Prudhomme, G. Rees, K. Stahl, and L. M. Tallaksen (2011), Large-scale river flow archives: Importance, current status and future needs, *Hydrol. Processes*, *25*(7), 1191–1200.
Hansen, M. C., et al. (2013), High-resolution global maps of 21st-century forest cover change, *Science*, *342*(6160), 850–853.
Hargreaves, G. L., G. H. Hargreaves, and J. P. Riley (1985), Irrigation water requirements for Senegal river basin, *J. Irrig. Drain. Eng.*, *111*(3), 265–275.
He, Y., A. Bárdossy, and E. Zehe (2011), A review of regionalisation for continuous streamflow simulation, *Hydrol. Earth Syst. Sci.*, *15*, 3539–3553.
Hengl, T., et al. (2014), SoilGrids1 km: Global soil information based on automated mapping, *PLOS One*, *9*(8), e105992.
Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis (2005), Very high resolution interpolated climate surfaces for global land areas, *Int. J. Climatol.*, *25*(15), 1965–1978.
Holwerda, F., L. A. Bruijnzeel, F. N. Scatena, H. F. Vugts, and A. G. C. A. Meesters (2012), Wet canopy evaporation from a Puerto Rican lower montane rain forest: The importance of realistically estimated aerodynamic conductance, *J. Hydrol.*, *414–415*, 1–15.
Hopmans, J. W., D. R. Nielsen, and K. L. Bristow (2002), How useful are small-scale soil hydraulic property measurements for large-scale vadose zone modeling?, in *Environmental Mechanics: Water, Mass and Energy Transfer in the Biosphere: The Philip Volume*, edited by D. Smiles, P. A. C. Raats, and A. Warrick, AGU, Washington, D. C., doi:10.1029/129GM20.
Hrachowitz, M., et al. (2013), A decade of predictions in ungauged basins (PUB): A review, *Hydrol. Sci. J.*, *58*(6), 1198–1255.
Hundecha, Y., and A. Bárdossy (2004), Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *J. Hydrol.*, *292*(1–4), 281–295.
Institute of Hydrology (1980), Low flow studies, *Tech. Rep. 1*, Wallingford, U. K.
Jain, S. K., and K. P. Sudheer (2008), Fitting of hydrologic models: A close look at the Nash-Sutcliffe index, *J. Hydrol. Eng.*, *13*(10), 981–986.
Jin, X., C. Xu, Q. Zhang, and Y. D. Chen (2009), Regionalization study of a conceptual hydrological model in Dongjiang basin, south China, *Quat. Int.*, *208*(1–2), 129–137.
Kauffeldt, A. (2014), Disinformative and uncertain data in global hydrology: Challenges for modelling and regionalisation, PhD thesis, Dep. of Earth Sci., Uppsala Univ., Swed.
Kauffeldt, A., S. Halldin, A. Rodhe, C.-Y. Xu, and I. K. Westerberg (2013), Disinformative data in large-scale hydrological modelling, *Hydrol. Earth Syst. Sci.*, *17*(7), 2845–2013.
Kay, A. L., D. A. Jones, S. M. Crooks, A. Calver, and N. S. Reynard (2006), A comparison of three approaches to spatial generalization of rainfall-runoff models, *Hydrol. Processes*, *20*(18), 3953–3973.
Kim, U., and J. J. Kaluarachchi (2008), Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia, *J. Hydrol.*, *362*(1–2), 39–56.
Kling, H., M. Fuchs, and M. Paulin (2012), Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, *J. Hydrol.*, *424–425*, 264–277, doi:10.1016/j.hydrol.2012.01.011.
Kokkonen, T. S., A. J. Jakeman, P. C. Young, and H. J. Koivusalo (2003), Predicting daily flows in ungauged catchments: Model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina, *Hydrol. Processes*, *17*(11), 2219–2238.
Lange, J. (2005), Dynamics of transmission losses in a large arid stream channel, *J. Hydrol.*, *306*(1–4), 112–126.
Lehner, B. (2012), Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network, *Tech. Rep. 41*, Global Runoff Data Cent., Fed. Inst. of Hydrol. (BfG), Koblenz, Germany.
Lehner, B., et al. (2011), High resolution mapping of the world's reservoirs and dams for sustainable river flow management, *Front. Ecol. Environ.*, *9*(9), 494–502.

Li, H., Y. Zhang, F. H. S. Chiew, and S. Xu (2009), Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index, *J. Hydrol.*, *370*(1–4), 155–162.
Li, H., M. Huang, M. S. Wigmosta, Y. Ke, A. M. Coleman, L. R. Leung, A. Wang, and D. M. Ricciuto (2011), Evaluating runoff simulations from the Community Land Model 4.0 using observations from flux towers and a mountainous watershed, *J. Geophys. Res.*, *116*, D24120, doi: 10.1029/2011JD016276.
Lidén, R., and J. Harlin (2000), Analysis of conceptual rainfall-runoff modelling performance in different climates, *J. Hydrol.*, *238*(3–4), 231–247.
Livneh, B., and D. P. Lettenmaier (2013), Regional parameter estimation for the unified land model, *Water Resour. Res.*, *19*, 100–114, doi: 10.1029/2012WR012220.
Lohmann, D., et al. (2004), Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project, *J. Geophys. Res.*, *109*, D07S91, doi:10.1029/2003JD003517.
Maier, H., et al. (2014), Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions, *Environ. Modell. Software*, *62*, 271–299.
Masih, I., S. Uhlenbrook, S. Maskey, and M. D. Ahmad (2010), Regionalization of a conceptual rainfall-runoff model based on similarity of the flow duration curve: A case study from the semi-arid Karkheh basin, Iran, *J. Hydrol.*, *391*(1–2), 188–201.
McDonnell, J. J., et al. (2007), Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology, *Water Resour. Res.*, *43*, W07301, doi:10.1029/2006WR005467.
McIntyre, N. R., H. Lee, H. S. Wheater, and A. R. Young (2004), Tools and approaches for evaluating uncertainty in streamflow predictions in ungauged UK catchments, in *Complexity and Integrated Resources Management, Proceedings of the IEMSS International Congress*, edited by C. Pahl-Wostl, et al., International Environmental Modelling and Software Society (iEMSs), Osnabrueck, Germany.
McIntyre, N. R., H. Lee, H. S. Wheater, A. Young, and T. Wagener (2005), Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, *41*, W12434, doi:10.1029/2005WR004289.
Mendoza, P. A., M. P. Clark, M. Barlage, B. Rajagopalan, L. Samaniego, G. Abramowitz, and H. Gupta (2015), Are we unnecessarily constraining the agility of complex process-based models?, *Water Resour. Res.*, *51*, 716–728, doi:10.1002/2014WR015820.
Merz, R., and G. Blöschl (2004), Regionalisation of catchment model parameters, *J. Hydrol.*, *287*(1–4), 95–123.
Merz, R., and G. Blöschl (2005), Flood frequency regionalisation: Spatial proximity vs. catchment attributes, *J. Hydrol.*, *302*(1–4), 283–306.
Minville, M., D. Cartier, C. Guay, L.-A. Leclaire, C. Audet, S. Le Digabel, and J. Merleau (2014), Improving process representation in conceptual hydrological model calibration using climate simulations, *Water Resour. Res.*, *50*, 5044–5073, doi:10.1002/2013WR013857.
Montanari, A. (2005), Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, *Water Resources Research*, *41*, W08406, doi:10.1029/2004WR003826.
Muller, J. P., G. López, G. Watson, N. Shane, T. Kennedy, P. Yuen, and P. Lewis (2011), The ESA GlobAlbedo project for mapping the Earth's land surface albedo for 15 years from European sensors, *Geophys. Res. Abstr.*, *13*, EGU2011-10969.
Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I—a discussion of principles, *J. Hydrol.*, *10*(3), 282–290.
Nasonova, O. N., Y. M. Gusev, and Y. E. Kovalev (2009), Investigating the ability of a land surface model to simulate streamflow with the accuracy of hydrological models: A case study using MOPEX materials, *J. Hydrometeorol.*, *10*(5), 1128–1150.
Nepstad, D. C., et al. (1994), The role of deep roots in the hydrological and carbon cycles of Amazonian forests and pastures, *Nature*, *372*(6507), 666–669.
Newman, A. J., et al. (2015), Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance, *Hydrol. Earth Syst. Sci.*, *19*, 209–223, doi: 10.5194/hess-19-209-2015.
Nijssen, B., G. M. O'Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood (2001), Predicting the discharge of global rivers, *J. Clim.*, *14*(15), 3307–3323.
Nijzink, R. C., L. Samaniego, J. Mai, R. Kumar, S. Thober, M. Zink, D. Schäfer, H. H. G. Savenije, and M. Hrachowitz (2015), The importance of topography controlled sub-grid process heterogeneity in distributed hydrological models, *Hydrol. Earth Syst. Sci. Discuss.*, *12*, 13,301–13,358, doi:10.5194/hessd-12-13301-2015.
Niu, G.-Y., et al. (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, *J. Geophys. Res.*, *116*, D12109, doi:10.1029/2010JD015139.
Oleson, K. W., D. M. Lawrence, G. B. Bonan, M. G. Flanner, E. Kluzek, P. J. Lawrence, S. Levis, S. C. Swenson, and P. E. Thornton (2010), *Technical description of version 4.0 of the Community Land Model (CLM), technical report NCAR/TN-478+STR*, Clim. and Global Dyn. Div., Natl. Cent. for Atmos. Res., Boulder, Colo.
Olson, D. M., et al. (2001), Terrestrial ecoregions of the world: A new map of life on Earth, *BioScience*, *51*(11), 933–938.
Osuch, M., R. J. Romanowicz, and M. J. Booij (2015), The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics, *Hydrol. Sci. J.*, *60*(7–8), 1299–1316.
Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine (2008), Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, *44*, W03413, doi:10.1029/2007WR006240.
Oudin, L., A. Kay, V. Andréassian, and C. Perrin (2010), Are seemingly physically similar catchments truly hydrologically similar?, *Water Resour. Res.*, *46*, W11558, doi:10.1029/2009WR008887.
Parajka, J., R. Merz, and G. Blöschl (2005), A comparison of regionalisation methods for catchment model parameters, *Hydrol. Earth Syst. Sci.*, *9*(3), 157–171.
Parajka, J., G. Blöschl, and R. Merz (2007), Regional calibration of catchment models: Potential for ungauged catchments, *Water Resour. Res.*, *43*, W06406, doi:10.1029/2006WR005271.
Parajka, J., A. Viglione, M. Rogger, J. L. Salinas, M. Sivapalan, and G. Blöschl (2013), Comparative assessment of predictions in ungauged basins: Part 1: Runoff-hydrograph studies, *Hydrol. Earth Syst. Sci.*, *17*, 1783–1795.
Patil, S. D., and M. Stieglitz (2014), Modeling daily streamflow at ungauged catchments: What information is necessary?, *Hydrol. Processes*, *28*(3), 1159–1169.
Peña-Arancibia, J. L., A. I. J. M. Van Dijk, M. Mulligan, and L. A. Bruijnzeel (2010), The role of climatic and terrain attributes in estimating baseflow recession in tropical catchments, *Hydrol. Earth Syst. Sci.*, *14*(11), 2193–2205.
Peel, M. C., F. H. S. Chiew, A. W. Western, and T. A. McMahon (2000), Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments, NCAR/TN-478+STR report prepared for the Australian National Land and Water Resources Audit. Cent. for Environ. Appl. Hydrol., Univ. of Melbourne, Aust.

Penman, H. L. (1948), Natural evaporation from open water, bare soil and grass, *Proc. R. Soc. A*, *193*, 120–146.
Petheram, C., P. Rustomji, F. H. S. Chiew, and J. Vleeshouwer (2012), Rainfall-runoff modelling in northern Australia: A guide to modelling strategies in the tropics, *J. Hydrol.*, *462–463*, 28–41.
Pilgrim, D. H., T. G. Chapman, and D. G. Doran (1988), Problems of rainfall-runoff modelling in arid and semiarid regions, *Hydrol. Sci. J.*, *33*(4), 379–400.
Plesca, I., E. Timbe, J. F. Exbrayat, D. Windhorst, P. Kraft, P. Crespo, K. B. Vachéa, H. G. Frede, and L. Breuer (2012), Model intercomparison to explore catchment functioning: Results from a remote montane tropical rainforest, *Ecol. Modell.*, *239*, 3–13.
Rakovec, O., et al. (2016), Multiscale and multivariate evaluation of water fluxes and states over European river basins, *J. Hydrometeorol.*, *17*(1), 287–307.
Razavi, T., and P. Coulibaly (2013), Streamflow prediction in ungauged basins: Review of regionalization methods, *J. Hydrol. Eng.*, *18*(8), 958–975.
Reichl, J. P. C., A. W. Western, N. R. McIntyre, and F. H. S. Chiew (2009), Optimization of a similarity measure for estimating ungauged streamflow, *Water Resour. Res.*, *45*, W10423, doi:10.1029/2008WR007248.
Rosero, E., L. E. Gulden, and Z. Yang (2011), Ensemble evaluation of hydrologically enhanced Noah-LSM: Partitioning of the water balance in high-resolution simulations over the Little Washita River experimental watershed, *J. Hydrometeorol.*, *12*(1), 45–64.
Samaniego, L., R. Kumar, and S. Attinger (2010), Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resour. Res.*, *46*, W05523, doi:10.1029/2008WR007327.
Samuel, J., P. Coulibaly, and R. Metcalfe (2011), Estimation of continuous streamflow in Ontario ungauged basins: Comparison of regionalization methods, *J. Hydrol. Eng.*, *16*(5), 447–459.
Schaefli, B., and H. V. Gupta (2007), Do Nash values have value?, *Hydrol. Processes*, *21*(15), 2075–2080.
Seibert, J. (1999), Regionalisation of parameters for a conceptual rainfall-runoff model, *Agric. For. Meteorol.*, *98–99*(1–4), 279–293.
Seibert, J., and M. J. P. Vis (2012), Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, *Hydrol. Earth Syst. Sci.*, *16*(9), 3315–3325.
Sellami, H., I. La Jeunesse, S. Benabdallah, N. Baghdadi, and M. Vanclooster (2014), Uncertainty analysis in model parameters regionalization: A case study involving the SWAT model in Mediterranean catchments (Southern France), *Hydrol. Earth Syst. Sci.*, *18*, 2393–2413, doi:10.5194/hess-18-2393-2014.
Shafii, M., and B. A. Tolson (2015), Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives, *Water Resour. Res.*, *51*, 3796–3814, doi:10.1002/2014WR016520.
Shu, C., and D. H. Burn (2003), Spatial patterns of homoheneous pooling groups for flood frequency analysis, *Hydrol. Sci. J.*, *48*(4), 601–618.
Shuttleworth, W. (1993), *Handbook of Hydrology, chap. 4, Evaporation*, McGraw-Hill, N. Y.
Siebert, S., P. Döll, J. Hoogeveen, J. Faures, K. Frenken, and S. Feick (2005), Development and validation of the global map of irrigation areas, *Hydrol. Earth Syst. Sci.*, *9*, 535–547, doi:10.5194/hess-9-535-2005.
Singh, R., S. A. Archfield, and T. Wagener (2014), Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments: A comparative hydrology approach, *J. Hydrol.*, *517*, 985–996, doi:10.1016/j.hydrol.2014.06.030.
Sivapalan, M. (2003), Prediction in ungauged basins: A grand challenge for theoretical hydrology, *Hydrol. Processes*, *17*(15), 3163–3170.
Slater, A. G., T. J. Bohn, J. L. McCreight, M. C. Serreze, and D. P. Lettenmaier (2007), A multimodel simulation of pan-Arctic hydrology, *J. Geophys. Res.*, *112*, G04S45, doi:10.1029/2006JG000303.
Sooda, A., and V. Smakhtin (2015), Global hydrological models: A review, *Hydrol. Sci. J.*, *60*(4), 549–565, doi:10.1016/j.jhydrol.2012.09.002.
Steele-Dunne, S., P. Lynch, R. McGrath, T. Semmler, S. Wang, J. Hanafin, and P. Nolan (2008), The impacts of climate change on hydrology in Ireland, *J. Hydrol.*, *356*(1–2), 28–45.
Stewart, I. T., D. R. Cayan, and M. D. Dettinger (2005), Changes toward earlier streamflow timing across western North America, *J. Clim.*, *18*(8), 1136–1155.
Stillman, S., X. Zeng, and M. G. Bosilovich (2016), Evaluation of 22 precipitation and 23 soil moisture products over a semiarid area in southeastern Arizona, *J. Hydrometeorol.*, *17*(1), 211–230.
Tait, A., R. Henderson, R. Turner, and X. Zheng (2006), Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface, *Int. J. Climatol.*, *26*(14), 2097–2115.
Tanaka, N., T. Kume, N. Yoshifuji, K. Tanaka, H. Takizawa, K. Shiraki, C. Tantasirin, N. Tangtham, and M. Suzuki (2008), A review of evapotranspiration estimates from tropical forests in Thailand and adjacent regions, *Agric. For. Meteorol.*, *148*(5), 807–819.
Te Linde, A. H., J. C. J. H. Aerts, R. T. W. L. Hurkmans, and M. Eberle (2008), Comparing model performance of two rainfall-runoff models in the Rhine Basin using different atmospheric forcing data sets, *Hydrol. Earth Syst. Sci.*, *12*(3), 943–957.
Trambauer, P., S. Maskeya, H. Winsemius, M. Werner, and S. Uhlenbrook (2013), A review of continental scale hydrological models and their suitability for drought forecasting in (sub-Saharan) Africa, *Phys. Chem. Earth*, *66*, 16–26.
Troch, P. A., G. Carrillo, M. Sivapalan, T. Wagener, and K. Sawicz (2013), Climate-vegetation-soil interactions and long-term hydrologic partitioning: Signatures of catchment co-evolution, *Hydrol. Earth Syst. Sci.*, *17*, 2209–2217, doi:10.5194/hess-17-2209-2013.
Troy, T. J., E. F. Wood, and J. Sheffield (2008), An efficient calibration method for continental-scale land surface modeling, *Water Resour. Res.*, *44*, W09411, doi:10.1029/2007WR006513.
Van Beek, L. P. H., and M. F. P. Bierkens (2009), The global hydrological model PCR-GLOBWB: Conceptualization, parameterization and verification, technical report, Utrecht Univ. [Available at http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf.]
Van Dijk, A. I. J. M. (2010), Climate and terrain factors explaining streamflow response and recession in Australian catchments, *Hydrol. Earth Syst. Sci.*, *14*(1), 159–169.
Van Dijk, A. I. J. M., J. L. Peña-Arancibia, E. F. Wood, J. Sheffield, and H. E. Beck (2013), Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide, *Water Resour. Res.*, *49*, 2729–2746, doi:10.1002/wrcr.20251.
Vandewiele, G. L., and A. Elias (1995), Monthly water balance of ungauged catchments obtained by geographical regionalization, *J. Hydrol.*, *170*(1–4), 277–291.
Vetter, T., S. Huang, V. Aich, T. Yang, X. Wang, V. Krysanova, and F. Hattermann (2015), Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents, *Earth Syst. Dyn.*, *6*(1), 17–43.
Viney, N. R., J. Vaze, F. H. S. Chiew, J. Perraud, D. A. Post, and J. Teng (2009), Comparison of multi-model and multi-donor ensembles for regionalisation of runoff generation using five lumped rainfall-runoff models, in *18th World IMACS/MODSIM Congress*, edited by R. S. Anderssen, R. D. Braddock and L. T. H. Newham, Modelling and Simulation Society of Australia and New Zealand Inc. Cairns, Aust.
Vis, M., R. Knight, S. Pool, W. Wolfe, and J. Seibert (2015), Model calibration criteria for estimating ecological flow characteristics, *Water*, *7*(5), 2358–2381.

Wagener, T., and H. S. Wheater (2006), Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, *320*(1–2), 132–154.
Wagener, T., M. Sivapalan, P. A. Troch, and R. Woods (2007), Catchment classification and hydrologic similarity, *Geogr. Compass*, *1*(4), 901–931.
Wallner, M., U. Haberlandt, and J. Dietrich (2013), A one-step similarity approach for the regionalization of hydrological model parameters based on self-organizing maps, *J. Hydrol.*, *494*, 59–71, doi:10.1016/j.jhydrol.2013.04.022.
Wanders, N., M. F. P. Bierkens, S. M. de Jong, and A. de Roo (2014), The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models, *Water Resour. Res.*, *50*, 6874–6891, doi:10.1002/2013WR014639.
Wang, Q. J. (1997), Using genetic algorithms to optimise model parameters, *Environ. Modell. Software*, *12*(1), 27–34.
Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo (2014), The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resour. Res.*, *50*, 7505–7514, doi:10.1002/2014WR015638.
Widén-Nilsson, E., S. Halldin, and C. Xua (2007), Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation, *J. Hydrol.*, *340*(1–2), 105–118.
Winsemius, H. C., B. Schaefli, A. Montanari, and H. H. G. Savenije (2009), On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, *Water Resour. Res.*, *45*, W12422, doi:10.1029/2009WR007706.
Wohl, E., et al. (2012), The hydrology of the humid tropics, *Nat. Clim. Change*, *2*, 655–662, doi:10.1038/nclimate1556.
Xia, Y., et al. (2012), Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow, *J. Geophys. Res.*, *117*, D03110, doi:10.1029/2011JD016048.
Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu (2007), A gauge-based analysis of daily precipitation over East Asia, *J. Hydrometeorol.*, *8*(3), 607–626.
Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, *30*, 1756–1774, doi:10.1016/j.advwatres.2007.01.005.
Ye, W., B. C. Bates, N. R. Viney, M. Sivapalan, and A. J. Jakeman (1997), Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments, *Water Resour. Res.*, *33*(1), 153–166.
Yokoo, Y., S. Kazama, M. Sawamoto, and H. Nishimura (2001), Regionalization of lumped water balance model parameters based on multiple regression, *J. Hydrol.*, *246*(1–4), 209–222.
Young, A. R. (2006), Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model, *J. Hydrol.*, *320*(1–2), 155–172.
Zaitchik, B. F., M. Rodell, and F. Olivera (2010), Evaluation of the global land data assimilation system using global river discharge data and a source-to-sink routing scheme, *Water Resour. Res.*, *46*, W06507, doi:10.1029/2009WR007811.
Zhang, L., N. Potter, K. Hickel, Y. Zhang, and Q. Shao (2008), Water balance modeling over variable time scales based on the Budyko framework – model development and testing, *J. Hydrol.*, *360*(1–4), 117–131.
Zhang, X., and G. Lindström (1996), A comparative study of a Swedish and a Chinese hydrological model, *J. Am. Water Resour. Assoc.*, *32*(5), 985–994.
Zhang, Y., and F. H. S. Chiew (2009), Relative merits of different methods for runoff predictions in ungauged catchments, *Water Resour. Res.*, *45*, W07412, doi:10.1029/2008WR007504.
Zhang, Y., J. Vaze, F. H. S. Chiew, and M. Li (2015), Comparing flow duration curve and rainfall-runoff modelling for predicting daily runoff in ungauged catchments, *J. Hydrol.*, *525*, 72–86.
Zhou, X., Y. Zhang, Y. Wang, H. Zhang, J. Vaze, L. Zhang, Y. Yang, and Y. Zhou (2012), Benchmarking global land surface models against the observed mean annual runoff from 150 large basins, *J. Hydrol.*, *470–471*, 269–279.